Inférence statistique des modèles conditionnellement hétéroscédastiques avec innovations stables, contraste non gaussien et volatilité mal spécifiée

Guillaume Lepage

Année : 2012

THÈSE

présentée à

L'UNIVERSITÉ Charles de Gaulle - Lille III

pour l'obtention du titre de

DOCTEUR EN MATHÉMATIQUES APPLIQUÉES

présentée et soutenue par

Guillaume LEPAGE

Inférence statistique des modèles conditionnellement hétéroscédastiques avec innovations stables, contraste non gaussien et volatilité mal spécifiée

Jury

Christian FRANCQ	Université Lille 3	Directeur de thèse
Jean-Michel ZAKOÏAN	Université Lille 3	Directeur de thèse
Siegfried HÖRMANN	Université Libre de Bruxelles	Rapporteur
Shiqing LING	Hong Kong University of Science and Technology	Rapporteur
Serge DAROLLES	Université Paris–Dauphine	Examinateur
Thomas SIMON	Université Lille 1	Examinateur
Olivier WINTENBERGER	Université Paris–Dauphine	Examinateur

Remerciements

À venir.

Résumé. Dans cette thèse, nous nous intéressons à l'estimation de modèles conditionnellement hétéroscédastiques sous différentes hypothèses. La thèse comporte trois parties et un chapitre introductif. Dans une première partie, en modifiant l'hypothèse d'identification usuelle du modèle, nous définissions un estimateur de quasi maximum de vraisemblance non gaussien et nous montrons que, sous certaines conditions, cet estimateur est plus efficace que l'estimateur du quasi maximum de vraisemblance gaussien. Nous étudions dans une deuxième partie l'inférence d'un modèle conditionnellement hétéroscédastique dans le cas où le processus des innovations est distribué selon une loi alpha stable. Nous établissons la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance. La loi alpha stable n'apparaissant que comme loi limite, nous étudions ensuite le comportement de ce même estimateur dans le cas où la loi du processus des innovations n'est plus une loi alpha stable mais est dans le domaine d'attraction d'une telle loi. Dans la dernière partie de cette thèse, nous étudions l'estimation d'un modèle GARCH lorsque le processus générateur de données est un modèle conditionnellement hétéroscédastique dont les coefficients sont sujets à des changements de régimes markoviens. Nous montrons que cet estimateur, dans un cadre mal spécifié, converge vers une pseudo vraie valeur et nous établissons sa loi asymptotique. Nous étudions cet estimateur lorsque le processus observé est stationnaire mais nous détaillons également ses propriétés asymptotiques lorsque ce processus est non stationnaire et explosif. Par des simulations, nous étudions les capacités prédictives du modèle GARCH mal spécifié. Nous déterminons ainsi la robustesse de ce modèle et de l'estimateur du quasi maximum de vraisemblance à une erreur de spécification de la volatilité.

Mots clés. Modèles conditionnellement hétéroscédastiques ; Modèles GARCH ; Quasi Maximum de Vraisemblance ; Estimateur efficace ; Lois alpha stables ; Domaine d'attraction ; Mesures de risques ; Value-at-Risk ; Modèles mal spécifiés ; Modèles à changement de régimes Markoviens ; Modèles non stationnaires.

Abstract. In this thesis, we focus on the inference of conditionally heteroskedastic models under different assumptions. This thesis consists of three parts and an introductory chapter. In the first part, we use an alternate identification assumption of the model and we define a non Gaussian quasi maximum likelihood estimator. We show that, under certain conditions, this estimator is more efficient than the Gaussian quasi maximum likelihood estimator. In a second part, we study the inference of a conditionally heteroskedastic model when the process of the innovations is distributed as an alpha stable law. We establish the consistency and the asymptotic normality of the maximum likelihood estimator. Since the alpha stable laws appear in general as a limit, we then focus of the behavior of this same estimator when the law of the innovation process is not stable but in the domain of attraction of a stable law. In the last part of this thesis, we study the estimation of a GARCH model when the data generating process is a conditionally heteroskedastic model whose coefficients are subject to Markov switching regimes. We show that, in a misspecified framework, this estimator converges toward a pseudo true value and we establish its asymptotic normality. We study this estimator when the observed process is stationary but we also give its asymptotic properties when this process is non stationary and explosive. Through simulations, we investigate the predictive ability of the misspecified GARCH model. Thus we determine the robustness of the model and of the estimator of the quasi maximum likelihood to the misspecification of the volatility.

Keywords. Conditionally heteroskedastic models ; GARCH models ; Quasi Maximum Likelihood ; Efficient estimator ; Alpha stable laws ; Domain of attraction ; Risk Measures ; Value-at-Risk ; Misspecified models ; Regime–switching models ; Non stationarity.

Table des matières

Chapitre 1

Introduction

Cette thèse porte essentiellement sur l'inférence des modèles de type ARCH (autorégressifs conditionnellement hétéroscédastiques). Les modèles GARCH (ARCH généralisés) et leurs nombreuses extensions ont été introduits pour tenir compte de caractéristiques communes aux séries financières de rendements.

Dans cette introduction, nous commençons par présenter ces caractéristiques communes, puis nous décrivons comment les modèles GARCH permettent d'en rendre compte, au moins partiellement. Nous synthétisons ensuite les résultats obtenus dans cette thèse, chapitre par chapitre, en les confrontant à la littérature existante et en insistant sur les motivations et les implications pratiques. Nous présentons parfois les résultats de manière informelle, en reportant le détail des énoncés, et bien entendu les démonstrations, aux chapitres suivants.

Nous nous intéressons successivement à l'estimation des GARCH :

(i) sans avoir à spécifier la loi des innovations

(ii) en présence d'une éventuelle erreur de spécification de la loi des innovations

(iii) en présence d'une éventuelle erreur de spécification du modèle.

Lorsque le modèle GARCH est supposé bien spécifié, et que seuls les paramètres de ce modèle sont intéressants, il semble que le cadre (i) soit le mieux adapté. L'estimateur usuel des modèles GARCH, le Quasi Maximum de Vraisemblance (QMV) gaussien se place dans ce cadre. Nous verrons qu'il est possible de battre le QMV gaussien en termes de précision, tout en restant dans le cadre (i) où on évite des hypothèses fortes sur la distribution des innovations. Il est parfois des situations où il convient de spécifier la loi des innovations. Pour certaines applications financières, en particulier pour le calcul de ce que l'on nomme « valeur à risque » (VaR) conditionnelle, il est utile de spécifier des quantiles, ou d'autres caractéristiques, de la loi des innovations. C'est également le cas lorsque les innovations sont à queues très épaisses, notamment les lois alpha stables, pour lesquelles le QMV gaussien ne fonctionne pas. Dans ce cadre, la méthode de choix est le maximum de vraisemblance (MV). D'un point de vue pratique, on s'expose alors aux deux types d'erreurs de spécification mentionnés dans les points (ii) et (iii). Plus précisément nous considérons le maximum de vraisemblance sous l'hypothèse que la loi des erreurs est alpha stable. La classe des lois alpha stables regroupe l'ensemble des lois limites du théorème central limite, que l'on soit dans un cadre classique où la variance existe (la loi limite est alors la gaussienne, qui est une loi stable particulière de paramètre $\alpha = 2$) ou dans le cas de variables à queues plus épaisses. Les séries financières étant réputées avoir des queues de distribution très lourdes, il semble pertinent d'autoriser des erreurs ayant une loi stable non nécessairement gaussienne. La loi alpha stable n'apparaît cependant que comme loi limite. Il est donc tout à fait plausible que la loi des innovations ne soit pas alpha stable, mais soit seulement dans le domaine d'attraction d'une loi alpha stable. Nous considérons ce type d'erreur de spécification, et nous cherchons à déterminer si la robustesse du QMV gaussien à une erreur de la forme (ii) dans le cadre $\alpha = 2$ s'étend au cas α quelconque. Nous considérons ensuite une erreur de spécification de type (iii), en considérant que les coefficients du modèle GARCH sont potentiellement sujets à des changements de régime markoviens. De tels changements de dynamique peuvent correspondre à des phases d'expansions et de contractions de l'économie, ou encore à la survenue de crises économiques ou financières soudaines et inattendues. Il est intéressant d'étudier le comportement des modèles classiques lorsque surviennent de tels changements de régime. L'objet du dernier chapitre sera donc de déterminer si un modèle GARCH standard est très sensible à une erreur de type (iii), ou s'il jouit d'une certaine propriété de robustesse aux changements de régime. Nous terminons par nos conclusions et nos perspectives de recherche.

1.1 Séries financières et modèles de volatilité.

La modélisation des séries financières est un domaine qui ne cesse d'être étudié. Les séries des prix des actions ou des indices étant clairement non stationnaires, on s'est principalement intéressé aux rendements [1] de ces séries. On a longtemps majoritairement supposé que ce processus des rendements était un processus gaussien indépendant et identiquement distribué (i.i.d.), cette hypothèse remonte à la thèse de Bachelier (1900) dont les travaux anticipaient l'introduction du mouvement Brownien. Si le processus des rendements de séries financières peut souvent être vu comme un bruit blanc, c'est-à-dire une suite de variables centrées, de variance finie et non corrélées, on peut cependant exhiber de nombreuses régularités statistiques (ou faits stylisés) communes à ces séries. Dans les premières références exhibant ces propriétés et les vérifiant empiriquement, nous pouvons citer Mandelbrot (1963) et Fama (1965). On peut énumérer ces régularités de la façon suivante

(i) Les rendements de ces séries ne sont en général pas corrélés. Cette propriété n'est souvent plus vérifiée dans le cas de données à très haute fréquence où des effets de microstructure peuvent apparaître et causer des autocorrélations négatives, voir par exemple Lo and Craig MacKinlay (1990).

(ii) Les séries des carrés des rendements ou des valeurs absolues des rendements sont souvent fortement autocorrélées. Cette propriété n'est pas incompatible avec l'hypothèse supposant que la série est un bruit blanc mais démontre simplement que le bruit blanc n'est pas indépendant.

(iii) On trouve également dans ces séries un regroupement des fortes variations, que l'on appelle *volatility clustering* en anglais. Cela est lié avec le point précédent, une forte variation de prix est souvent suivie par une autre forte variation, dans un sens ou dans l'autre.

(iv) On observe également que ces séries possèdent des pics de distribution autour de 0 ainsi que des queues épaisses. Elles sont appelées leptokurtiques. En conséquence, certains moments d'ordre supérieur de ces séries peuvent ne pas exister.

Ces propriétés illustrent la difficulté de modéliser les séries financières. Ces séries présentent en effet une hétéroscédasticité conditionnelle : conditionnellement aux valeurs passées, la variance du processus $(\epsilon_t)_t$ (cette notation sera utilisée par la suite pour désigner la série des rendements étudiée) n'est pas constante. En effet, si les dernières valeurs prises par le processus $(\epsilon_t)_t$ sont de grandes ampleurs, alors la variance conditionnelle sera plus forte qu'après une période de faibles mouvements. On peut noter qu'un processus peut être conditionnellement hétéroscédastique sans l'être inconditionnellement [2]. On peut mathématiquement définir l'hétéroscédasticité de la façon suivante, si $(\epsilon_t)_t$ admet une variance, alors il est conditionnellement hétéroscédastique si

$$\mathrm{Var}\left[\epsilon_t \middle| \, \epsilon_{t-1}, \epsilon_{t-2}, \cdots\right] \neq \mathrm{cste}.$$

Du fait de cette propriété, la modélisation classique du type ARMA (pour *AutoRegressive Moving Average*) est ici inopérante. On définit donc des modèles conditionnellement hétéroscédas-

1. Le rendement r_t (ou log rendement) d'une série $(X_t)_t$ peut être défini par $r_t = \log\left(\frac{X_t}{X_{t-1}}\right)$, pour $t \geq 1$.

2. Un processus $(X_t)_t$ est inconditionnellement hétéroscédastique (ou hétéroscédastique) si sa dispersion n'est pas constante dans le temps. Dans le cas où ce processus admet une variance, $(X_t)_t$ est hétéroscédastique si $\mathrm{Var} X_t \neq \mathrm{Cste}$.

tiques, qui décomposent le processus observé $(\epsilon_t)_t$ de la façon suivante :

$$\epsilon_t = \sigma_t \eta_t, \tag{1.1.1}$$

où $(\eta_t)_t$ est un processus i.i.d. centré et de variance unité. Le processus positif $(\sigma_t)_t$ est appelé volatilité du processus $(\epsilon_t)_t$. Beaucoup de modèles ont été présentés dans la littérature pour expliquer cette hétéroscédasticité conditionnelle. On en énonce trois familles

- Les processus du type ARCH (*AutoRegressive Conditionaly Heteroskedastic*). Cette famille est très certainement la plus utilisée, la volatilité du processus y est spécifiée comme une fonction déterministe du passé du processus et pour la généralisation GARCH (*Generalized* ARCH) des valeurs passées de la volatilité.
- Les modèles à volatilité stochastique représentent une alternative aux modèles du type GARCH. Dans ces modèles, la volatilité est elle-même un processus latent indépendant du processus $(\eta_t)_t$, voir Taylor (2005) ou Ghysels et al. (1996).
- Les modèles à changement de régime stochastique pour lesquels la volatilité s'écrit comme une fonction des valeurs passées du processus observé ainsi que de la valeur prise par Δ_t, un processus latent indépendant du processus des innovations $(\eta_t)_t$. Le processus $(\Delta_t)_t$ est souvent modélisé par une chaîne de Markov à espace d'états finis.

On s'intéressera dans cette thèse aux modèles issus de la famille des GARCH. Si on appelle $\mathcal{F}_t$, la tribu engendrée par les valeurs passées du processus $(\epsilon_t)_t$, ces modèles vérifient $\sigma_t \in \mathcal{F}_t$, la volatilité est mesurable par la tribu engendrée par les valeurs passées du processus $(\epsilon_t)_t$. Si le processus $(\epsilon_t)_t$ admet une variance, on peut alors identifier σ_t^2 comme la variance conditionnelle de $(\epsilon_t)_t$. On a dans ce cas

$$\mathrm{Var}\left[\epsilon_t \mid \epsilon_{t-1}, \epsilon_{t-2}, \cdots\right] = \sigma_t^2$$

De très nombreux modèles appartenant à la famille des GARCH ont été présentés dans la littérature économétrique. On y trouve par exemple des spécifications permettant d'expliquer l'asymétrie que l'on trouve chez certaines séries financières, voir par exemple Nelson (1991), Zakoian (1994) ou Hörmann (2008). On peut trouver un glossaire détaillant tous les modèles proposés avant cette date chez Bollerslev (2008). Dans cette thèse, on s'intéressera à la famille des modèles GARCH, que l'on présente plus en détails dans la section suivante.

Les modèles de la famille GARCH permettent d'expliquer et donc de prévoir la volatilité des séries financières observées. Ces modèles sont donc particulièrement utiles pour le calcul et le contrôle des risques financiers. Une mesure de risques très répandue est la VaR (*Value at Risk*) de niveau α, elle est définie comme le montant des pertes qui ne devrait être dépassé qu'avec une probabilité α. Dans le cas d'une modélisation du type (1.1.1), la probabilité conditionnellement au passé que ϵ_t soit inférieur à un certain seuil peut s'exprimer comme le produit de la volatilité conditionnelle par un quantile de la distribution du processus $(\eta_t)_t$.

1.2 Modèle GARCH

Comme mentionné dans la section précédente, la classe des modèles GARCH est une classe extrêmement utilisée dans le cadre de la modélisation des séries financières. Le processus observé s'écrit comme dans l'équation (1.1.1). Dans le modèle ARCH, introduit par Engle (1982), on tient compte de la persistance de volatilité en faisant dépendre $(\sigma_t)_t$ des différentes valeurs passées de

ϵ_t^2. Ainsi, on écrit

$$\begin{cases} \epsilon_t = \sigma_t \eta_t, \quad (\eta_t)_t \text{ i.i.d } (0,1) \\ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2, \end{cases} \tag{1.2.1}$$

où $(\eta_t)_t$ i.i.d $(0,1)$ indique que le processus $(\eta_t)_t$ est i.i.d., centré et de variance unité. Ce type de modèle permet d'obtenir des autocorrélations non nulles pour ϵ_t^2, cependant on observe sur les séries financières une décroissance assez lente des autocorrélations. Une modélisation ARCH réaliste nécessiterait ainsi un ordre q très élevé. La généralisation GARCH de Bollerslev (1986) permet d'obtenir des autocorrélogrammes plus conformes à ce qui est observé sur les séries financières sans devoir estimer un grand nombre de paramètres. Par critère de parcimonie, ce modèle est donc souvent préféré aux modèles ARCH. Cette spécification s'écrit

$$\begin{cases} \epsilon_t = \sigma_t \eta_t, \quad (\eta_t)_t \text{ i.i.d } (0,1) \\ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \end{cases} \tag{1.2.2}$$

On peut noter que cette écriture fait dépendre σ_t des valeurs passées de ϵ_t^2 et de ses propres valeurs passées. Sous certaines conditions d'inversibilité, on obtient une écriture faisant dépendre $(\sigma_t)_t$ uniquement des valeurs passées de $(\epsilon_t^2)_t$ et ainsi vérifier que $\sigma_t \in \mathcal{F}_t$. Il est bien sûr nécessaire d'assurer la positivité de σ_t, on impose généralement que le paramètre θ du modèle vérifie

$$\theta = (\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)' \in \mathbb{R}_+^* \times \mathbb{R}_+^{p+q}.$$

On note également que supposer que le processus des innovations $(\eta_t)_t$ admet une variance unité est nécessaire pour que le modèle soit identifiable. On verra par la suite que d'autres hypothèses d'identification peuvent être faites, l'hypothèses $E\eta_t^2 = 1$ possède l'avantage de permettre d'identifier σ_t^2 comme la volatilité conditionnelle de ϵ_t.

La première propriété mathématique qu'il convient d'obtenir pour étudier ce modèle est la stricte stationnarité[3] du processus $(\epsilon_t)_t$. Le cas GARCH$(1,1)$ a été traité par Nelson (1990), puis Bougerol and Picard (1992) ont obtenu une condition pour un modèle GARCH(p,q) général. L'idée est d'écrire le modèle sous forme vectorielle Markovienne. On introduit $h_t = \left(\epsilon_t^2, \ldots, \epsilon_{t-q+1}^2, \sigma_t^2, \ldots, \sigma_{t-p+1}^2\right)'$ et des matrices $\underline{b}_t$ et A_t telles que

$$\underline{h}_t = A_t \underline{h}_{t-1} + \underline{b}_t.$$

Avec de telles notations, on définit le coefficient de Lyapunov associé à la suite de matrices $(A_t)_t$ par

$$\gamma = \lim_{n \to +\infty} \frac{1}{t} E\left[\log \|A_t \cdots A_1\|\right]. \tag{1.2.3}$$

Une condition nécessaire et suffisante pour l'existence d'une solution stationnaire $(\epsilon_t)_t$ est $\gamma < 0$. Si cette condition est vérifiée, alors le processus solution $(\epsilon_t)_t$ est non seulement stationnaire mais aussi ergodique[4] et non anticipatif[5]. Si on veut prouver l'existence de moments pour le

3. Un processus $(X_t)_t$ est dit strictement stationnaire si les vecteurs $(X_1, \ldots, X_k)'$ et $(X_{1+h}, \ldots, X_{k+h})'$ ont même loi jointe, pour tout $k \in \mathbb{N}$ et pour tout $h \in \mathbb{N}$.

4. Informellement, cela signifie que la loi forte des grands nombres s'applique au processus.

5. Cela signifie que le processus est indépendant des innovations futures. ϵ_t ne dépend pas de η_k pour $k > t$.

processus $(\epsilon_t)_t$ ou la stationnarité au second-ordre de ce processus [6], on doit restreindre l'espace des paramètres et imposer

$$\sum_{i=1}^{q} \alpha_t + \sum_{j=1}^{p} \beta_j < 1.$$

Sans imposer cette condition, si $\gamma < 0$, on peut cependant prouver qu'il existe $s > 0$ tel que $E|\epsilon_t|^{2s} < +\infty$. Cette propriété est fondamentale pour prouver les résultats statistiques que nous énoncerons sous l'hypothèse de stationnarité stricte.

Une fois placés dans un cadre ergodique et stationnaire, nous pouvons nous intéresser à l'estimation des modèles GARCH. La méthode la plus employée est certainement la méthode du Quasi Maximum de Vraisemblance (QMV) utilisant la densité gaussienne. Cette méthode définit un estimateur maximisant la vraisemblance du modèle sous l'hypothèse que le processus des innovations est distribué selon une loi normale. Si on utilise la loi gaussienne pour définir l'estimateur, il n'est pas nécessaire de supposer pour autant que la vraie distribution du processus $(\eta_t)_t$ est gaussienne. C'est l'absence de cette hypothèse qui distingue le QMV de la méthode du Maximum de Vraisemblance (MV) gaussien. Les propriétés asymptotiques de cet estimateur ont été établies pour le modèle ARCH par Weiss (1986). Le cas GARCH(1,1) a été traité par Lumsdaine (1996), mais sous des conditions très restrictives sur les moments du processus des innovations ainsi que sur la forme de la distribution de ce processus. D'autres ont étudié cet estimateur sous une hypothèse de stationnarité au second ordre, voir Lee and Hansen (1994). Enfin, Berkes et al. (2003) et Francq and Zakoïan (2004) ont obtenu les propriétés asymptotiques de l'estimateur du QMV dans le cas GARCH(p,q) sous des hypothèses minimales. On peut également citer Straumann (2005) qui a énoncé des résultats similaires dans le cadre d'un modèle conditionnellement hétéroscédastique généralisé. Le modèle GARCH est parfois également utilisé comme bruit blanc d'un modèle ARMA, ce modèle est appelé ARMA-GARCH. Francq and Zakoïan (2004) ont étudié l'estimation par QMV d'un tel modèle sous une hypothèse d'existence d'un moment d'ordre 4 du processus $(\epsilon_t)_t$. Zhu and Ling (2011) utilisent un estimateur de quasi maximum de vraisemblance exponentiel pondéré et obtiennent la convergence ainsi que la normalité asymptotique de cet estimateur en affaiblissant cette hypothèse. Les propriétés asymptotiques du QMV dans le cas multivarié ont également été établies, voir par exemple Comte and Lieberman (2003) ou Bauwens et al. (2006).

L'estimateur du QMV présente de nombreux avantages : il ne requière que de très faibles hypothèses tant sur le processus des innovations (dont la distribution reste non spécifiée mais dont on doit supposer l'existence d'un moment d'ordre 4) que sur l'espace des paramètres et sur la vraie valeur des paramètres.

L'estimation par QMV gaussien du modèle (1.2.2) constituant la méthode de référence par rapport aux alternatives considérées dans cette thèse, on détaille maintenant sa définition et ses propriétés asymptotiques. Le critère à minimiser construit en supposant que le processus des innovations suit une loi gaussienne et donc que la loi conditionnelle de ϵ_t est une loi normale

6. Un processus $(X_t)_t$ est dit stationnaire au second-ordre si

(i) $EX_t^2 < +\infty \quad \forall t \in \mathbb{Z}$,

(ii) $EX_t = m \quad \forall t \in \mathbb{Z}$,

(iii) $\mathrm{Cov}(X_t, X_{t-h}) = \gamma_X(h) \quad \forall t, h \in \mathbb{Z}$.

La fonction γ_X est appelée fonction d'autocovariance de $(X_t)_t$.

centrée de variance σ_t^2, s'écrit

$$\tilde{I}_n = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta), \quad \text{où } \tilde{l}_t(\theta) = \log \tilde{\sigma}_t^2(\theta) + \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\theta)}. \tag{1.2.4}$$

Les $\tilde{\sigma}_t(\theta)$ sont définis de façon récurrente par,

$$\forall t \geq 1, \quad \tilde{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-i}^2.$$

On doit utiliser des valeurs initiales pour cette récurrence, on peut poser par exemple

$$\epsilon_0^2 = \cdots = \epsilon_{1-q}^2 = \tilde{\sigma}_0^2 = \cdots = \tilde{\sigma}_{1-p}^2 = \omega.$$

On définit également l'espace des paramètres $\Theta \subset \mathbb{R}_+^* \times \mathbb{R}_+^{p+q}$. Avec ces notations, on peut définir l'estimateur de quasi maximum de vraisemblance gaussien en posant

$$\theta_n = \underset{\theta\in\Theta}{\operatorname{argmin}}\ \tilde{I}_n(\theta). \tag{1.2.5}$$

Sous certaines hypothèses, cet estimateur est convergent et asymptotiquement normal, on a alors

$$\theta_n \to \theta_0, \quad \text{p.s.}$$

On a également

$$\sqrt{n}\,(\theta_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(E\eta_t^4 - 1\right) J^{-1}\right),$$

où J est défini par $J = E\left[\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right]$. La quantité $l_t(\theta_0)$ représente ici la limite stationnaire de $\tilde{l}_t(\theta_0)$ et sera définie plus formellement dans la suite de cette thèse.

De nombreux résultats sur la structure probabiliste de cette famille de modèles ont été et sont encore obtenus. Les propriétés de β mélange (β *mixing*) du modèle GARCH(p, q) ont été établies par Boussama (1998). En considérant le modèle GARCH comme un cas particulier d'un modèle markovien et en utilisant des résultats de Liebscher (2005), Meitz and Saikkonen (2008) ont pu étendre et généraliser ce résultat. On peut également citer Francq and Zakoian (2006) qui établissent les propriétés de mélanges pour une classe générale de modèles GARCH$(1, 1)$ et Fryzlewicz and Subba Rao (2011) qui étudient les propriétés de mélange de modèles ARCH(∞) et de modèles ARCH à coefficients dynamiques.

Basrak et al. (2002) ont également établi la propriété de variation régulière pour un processus GARCH(p, q). Les queues de distribution du processus $(\epsilon_t)_t$ ont un comportement de Pareto, c'est-à-dire que ce processus vérifie $\lim_{u\to+\infty} u^\kappa \mathbb{P}\left[\epsilon_t > u\right] = K$, où K est une constante. Des résultats similaires ainsi que des généralisations peuvent être trouvés parmi les références suivantes, Davis and Mikosch (2009a), Davis and Mikosch (2009b) et Mikosch et al. (2012).

1.3 Modèles mal spécifiés

Une partie de cette thèse est consacrée à l'estimation de modèles par la méthode du maximum de vraisemblance. Quand on fait l'hypothèse que le modèle est bien spécifié, l'estimateur du maximum de vraisemblance fournit en général une bonne estimation de la vraie valeur du paramètre. Dans la réalité, on sait que cette hypothèse est très forte, très restrictive. En effet, un modèle est en général utilisé car il permet d'expliquer certains faits stylisés vérifiés par la série étudiée mais il est utopique de penser que cette série a été générée par ce même modèle.

Il y a plusieurs façons de considérer un modèle mal spécifié. Dans le cas du modèle GARCH, la mauvaise spécification peut venir de la densité utilisée pour le processus des innovations $(\eta_t)_t$. Dans ce cas, on estime le vrai paramètre du modèle mais en optimisant une vraisemblance construite sous une hypothèse erronée. Il est alors intéressant d'étudier les propriétés asymptotiques de cet estimateur, de donner les conditions pour lesquelles il convergera vers le vrai paramètre de la spécification en dépit de l'erreur faite sur le choix de la loi du processus des innovations. Ce type de mauvaise spécification a été étudié par White (1982). C'est également dans ce type de problématique que l'on trouve l'origine de l'estimateur de quasi (ou pseudo) maximum de vraisemblance. Gourieroux et al. (1984) montrent que sous certaines conditions, l'estimateur du maximum de vraisemblance calculé sous l'hypothèse que la densité des innovations est gaussienne, reste convergent quand la vraie loi du processus des innovations appartient à une certaine famille de loi.

L'autre façon de mal spécifier un modèle est de calculer une vraisemblance à partir d'une mauvaise dynamique du modèle. On peut par exemple utiliser des ordres du modèle qui ne correspondent pas aux ordres du processus générateur des données (appelé par la suite DGP), par exemple estimer un modèle ARMA(2,1) alors que le processus est généré selon un ARMA(2,3). Il est alors intéressant d'étudier comment va se comporter l'estimateur mal spécifié ; va-t-il converger vers une pseudo vraie valeur qui maximiserait la "fausse" vraisemblance ? Le modèle estimé présente-t-il encore de l'intérêt dans le but de faire de la prévision sur la série ? Les références sur ce thème sont encore une fois à trouver chez Halbert White (voir Domowitz and White (1982); White (1984)). Dahlhaus and Wefelmeyer (1996) ont par exemple étudié le comportement d'un estimateur "Whittle" dans le cas d'un modèle ARMA mal spécifié. L'estimateur converge alors vers la valeur du paramètre qui minimise la distance de Kullback-Leibler entre les spécifications. Plus récemment on trouve également des travaux sur des modèles à chaînes de Markov cachées mal spécifiés, voir Douc and Moulines (2011). Du côté des modèles GARCH, on trouve plusieurs travaux étudiant l'estimation d'un modèle GARCH (souvent avec un estimateur "Whittle") sur un processus de diffusion. Ainsi Jensen and Lange (2010) prouvent que l'estimateur du quasi maximum de vraisemblance d'un GARCH(1,1) converge vers le jeu de paramètre $(0, 0, 1)$ quand la fréquence de discrétisation du processus augmente. De manière générale, on trouve plusieurs travaux remarquant que, quand le DGP n'est pas stationnaire (par exemple dans le cas d'un modèle avec des changements structurels) alors les paramètres estimés du GARCH(1,1) font apparaître un effet IGARCH et vérifient $\alpha + \beta = 1$. On peut ainsi lire Nelson (1992), Nelson and Foster (1995), Hillebrand (2005) ou Hu and Shin (2008). Il existe également des travaux mettant en œuvre des tests de spécification appliqués aux modèles GARCH permettant de détecter ces changements structurels. On peut citer par exemple Halunga and Orme (2010) et Lundbergh and Teräsvirta (2002). On peut également citer Aue et al. (2009) proposant de détecter des changements structurels pour des modèles multivariés.

1.4 Résultats du chapitre 2

L'objectif de ce chapitre est de proposer un estimateur du modèle (1.2.2) identifié sous la contrainte $E\eta_t^2 = 1$ qui peut être plus efficace que le QMLE qui servira de référence. Le sujet de ce chapitre est lié aux travaux de Berkes and Horváth (2004) qui ont également étudié l'estimation d'un modèle GARCH(p, q) utilisant une vraisemblance non gaussienne sous une reparamétrisation de ce modèle. On peut également citer Francq and Zakoian (2010) qui ont également utilisé un estimateur de QMV non gaussien pour améliorer la capacité de prédiction sur ces modèles. Ce chapitre se démarque de ces travaux par le fait que l'on y propose un estimateur du modèle identifié sous l'hypothèse classique $E\eta_t^2 = 1$ et que l'on y décrit une stratégie de test permettant de déterminer l'estimateur du QMV le plus efficace. Un estimateur non gaussien en deux étapes similaire a été depuis étudié par Qi et al. (2010).

1.4.1 Estimation du paramètre par QML non gaussien

Pour $r > 0$, on définit $\mu_r = E\left|\eta_t\right|^r$, en posant

$$\eta_t^{(r)} = \frac{1}{\mu_t^{1/r}}\eta_t, \quad \text{and } \sigma_t^{(r)} = \mu_t^{1/r}\sigma_t,$$

on peut écrire une version reparamétrée du modèle (1.2.2) qui s'écrit

$$\begin{cases} \epsilon_t = \sigma_{0t}^{(r)}\eta_t^{(r)} \\ \sigma_{0t}^{(r)2} = \omega_0^{(r)} + \sum_{i=1}^{q}\alpha_{0i}^{(r)}\epsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_{0j}^{(r)}\sigma_{0t-j}^{(r)2}, \end{cases} \tag{1.4.1}$$

La quantité $\theta_0^{(r)} = \left(\omega_0^{(r)}, \alpha_{01}^{(r)}, \cdots, \alpha_{0q}^{(r)}, \beta_{01}^{(r)}, \cdots, \beta_{0p}^{(r)}\right)'$ est alors la vraie valeur du paramètre sous cette reparamétrisation. En posant pour tout $s > 0$, $\mu_s^{(r)} = E\left|\eta_t^{(r)}\right|^s$ et en définissant la matrice

$$B^{(r)} = \begin{pmatrix} \mu_r^{-2/r} I_{q+1} & 0 \\ 0 & I_p \end{pmatrix} = \begin{pmatrix} \mu_2^{(r)} I_{q+1} & 0 \\ 0 & I_p \end{pmatrix},$$

on obtient facilement une relation entre θ_0 et $\theta_0^{(r)}$

$$\theta_0 = B^{(r)}\theta_0^{(r)}. \tag{1.4.2}$$

Le modèle d'origine est identifié par $E\eta_t^2 = 1$, le modèle reparamétré vérifie, lui $E\left|\eta_t^{(r)}\right|^r = 1$. Francq and Zakoian (2010) ont prouvé que cette condition d'identifiabilité impose de choisir une densité particulière pour que l'estimateur du QML soit consistant, il s'agit de la loi gaussienne généralisée que l'on peut écrire de la façon suivante, pour $r > 0$

$$h_r(x) = \frac{r^{1-1/r}}{2\Gamma(1/r)}\exp\left(-\frac{1}{r}|x|^r\right).$$

On écrit maintenant la vraisemblance du modèle (1.4.1) sous l'hypothèse que le processus $(\eta_t^{(r)})_t$ est distribué selon cette loi de probabilité. On pose alors

$$\tilde{\mathbf{I}}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta)$$

avec

$$\tilde{l}_t(\theta) = \log \tilde{\sigma}_t^2(\theta) + \frac{2}{r}\frac{|\epsilon_t|^r}{\tilde{\sigma}_t^r(\theta)} = -2\log\left(\frac{1}{\tilde{\sigma}_t(\theta)}h_r\left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}\right)\right) + K.$$

Les $\tilde{\sigma}_t$ sont définis par la récurrence du modèle (1.4.1) et en utilisant des valeurs initiales. Avec cette vraisemblance, on peut définir un estimateur de $\theta_0^{(r)}$ par

$$\widehat{\theta}_n^{(r)} = \underset{\theta\in\Theta^{(r)}}{\operatorname{argmin}}\, \tilde{\mathbf{I}}_n(\theta).$$

On a ainsi défini un estimateur du paramètre pour le modèle reparamétré (1.4.1). L'objectif était de définir un estimateur comparable à l'estimateur du QML gaussien, il faut donc corriger le biais causé par la reparamétrisation. Pour cela, on utilise la relation (1.4.2). En utilisant cette relation on n'a besoin que d'un estimateur de la matrice $B^{(r)}$ pour trouver un estimateur de θ_0. On définit ainsi

$$\hat{B}_n^{(r)} = \begin{pmatrix} \hat{\mu}_{2,n}^{(r)} I_{q+1} & 0 \\ 0 & I_p \end{pmatrix}, \text{ avec } \hat{\mu}_{2,n}^{(r)} = \frac{1}{n}\sum_{t=1}^{n}\left|\hat{\eta}_t^{(r)}\right|^s, \text{ et } \hat{\eta}_t = \frac{\epsilon_t}{\tilde{\sigma}_t\left(\widehat{\theta}_n^{(r)}\right)}.$$

On peut maintenant définir un estimateur de QML non gaussien en deux étapes pour le paramètre θ_0, en posant

$$\hat{\theta}_{n,r} = \hat{B}_n^{(r)}\hat{\theta}_n^{(r)}.$$

On appelle cette quantité estimateur du quasi maximum de vraisemblance en deux étapes ou 2QMLE.

On prouve maintenant sous certaines hypothèses que cet estimateur est convergent et asymptotiquement normal. Ces hypothèses sont pour la plupart celles nécessaires pour obtenir la convergence et la normalité asymptotique du QML gaussien. On a simplement besoin de supposer que le processus des innovations est tel que $|\eta_t|$ puisse prendre cinq valeurs différentes. Dans le cas où $r > 2$, on a également besoin de supposer que $(\eta_t)_t$ admet un moment d'ordre $2r$. Sous ces hypothèses (données dans le chapitre 2), on obtient le théorème suivant.

Théorème 1.4.1. Pour $r > 0$ et sous les hypothèses **A1-A4**, le 2QMLE θ_0 vérifie

$$\widehat{\theta}_{n,r} \underset{n\to+\infty}{\longrightarrow} \theta_0, \quad a.s.$$

Si, de plus, on a **A5-A6** alors

$$\sqrt{n}\left(\widehat{\theta}_{n,r} - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_r) \tag{1.4.3}$$

avec

$$\Sigma_r = g(r)J^{-1} + \{\mu_4 - 1 - g(r)\}\,\overline{\theta}_0\overline{\theta}_0', \qquad g(r) = \left(\frac{2}{r}\right)^2\left(\frac{\mu_{2r}}{\mu_r^2} - 1\right),$$

et $\overline{\theta}_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0q}, 0, \ldots, 0)'$, $J = E\left(\phi_t\phi_t'\right)$, $\phi_t = \phi_t(\theta_0)$, $\phi_t(\theta) = \frac{1}{\sigma_t^2(\theta)}\frac{\partial\sigma_t^2}{\partial\theta}(\theta)$.

Il est intéressant de remarquer que l'estimateur $\hat{\theta}_{n,2}$ correspond exactement à l'estimateur du QML gaussien dans le cas où on prend des valeurs nulles pour initialiser la récurrence des $(\tilde{\sigma}_t)_t$. Dans ce cas, le moment empirique $\hat{\mu}_{2,n}^{(2)}$ est exactement égal à 1.

L'efficacité asymptotique de l'estimateur $\hat{\theta}_{n,r}$ dépend alors de la fonction $g(r)$, l'efficacité augmente quand cette fonction décroit. On peut également prouver que $\hat{\theta}_{n,r}$ est plus efficace que $\hat{\theta}_{n,2}$ si la condition suivante est vérifiée

$$g(r) < \mu_4 - 1.$$

L'objectif est maintenant d'utiliser l'estimateur optimal, celui qui correspond à la valeur de $r > 0$ qui minimise la fonction g. On définit $R = [\underline{r}, \overline{r}] \subset (0, r_{\max})$ avec $r_{\max} = \sup\{r \in \mathbb{R} \; ; \; \mu_{2r} < \infty\}$ et on fait l'hypothèse suivante (hypothèse **A7** du chapitre 2)

A7 Il existe un unique $r_0 > 0$ tel que $r_0 = \underset{r \in R}{\operatorname{argmin}}\, g(r)$.

Cette hypothèse porte sur l'unicité de l'optimum r_0, l'existence peut être obtenue en utilisant la compacité de R et la continuité de la fonction g.

L'estimateur optimal est donc $\hat{\theta}_{n,r_0}$. La valeur de r_0 est en général inconnue, elle dépend de la distribution du processus $(\eta_t)_t$ qui n'est pas spécifiée, cet estimateur ne peut donc pas être utilisé en pratique. On peut cependant définir un estimateur de r_0 que l'on appellera $\hat{r}_n$ et utiliser l'estimateur $\hat{\theta}_{n,\hat{r}_n}$ qui sera alors parfaitement opérationnel.

1.4.2 Estimation et test de l'optimum r_0

La connaissance de r_0 permettrait d'obtenir l'estimateur le plus performant dans la famille des estimateurs du QML. L'objectif est maintenant de savoir si l'estimation par QML gaussien apporte une plus-value par rapport à l'estimateur du QML gaussien classiquement utilisé. La première étape consiste à estimer θ_0 par QML gaussien, puis on utilisera les résidus de cette estimation pour calculer $\hat{r}_n$ et déterminer si l'estimateur $\hat{\theta}_{r,\hat{r}_n}$ est asymptotiquement plus efficace que $\hat{\theta}_{n,2}$. On définit pour $u > 0$ tel que $\mu_u < +\infty$

$$\widehat{\mu}_{n,u} = \frac{1}{n} \sum_{t=1}^{n} |\widehat{\eta}_t|^u \quad \text{avec} \quad \widehat{\eta}_t = \frac{\epsilon_t}{\tilde{\sigma}_t\left(\widehat{\theta}_{n,2}\right)},$$

on définit également un estimateur de $g(r)$

$$\widehat{g}_n(r) = \left(\frac{2}{r}\right)^2 \left(\frac{\widehat{\mu}_{n,2r}}{\widehat{\mu}_{n,r}^2} - 1\right),$$

enfin, on peut définir un estimateur de r_0

$$\widehat{r}_n = \underset{r \in R}{\operatorname{argmin}}\, \widehat{g}_n(r).$$

Le théorème suivant donne les propriétés asymptotiques de $\hat{r}_n$.

Théorème 1.4.2. Sous les hypothèses **A1-A4** et **A7** (telles qu'énoncées dans le chapitre 2), et si $E\left\{|\eta_t|^{4r_0} (\log|\eta_t|)^2\right\} < \infty$, on a $\widehat{r}_n \to r_0$ a.s. et si de plus r_0 appartient à l'intérieur de R, alors

$$\sqrt{n}\left(\widehat{r}_n - r_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \tau_{r_0}\right)$$

où τ_{r_0} ne dépend que de la distribution de $(\eta_t)_t$.

On peut maintenant mettre un place un test statistique pour déterminer si l'estimateur du QML gaussien $\hat{\theta}_{n,2}$ est l'estimateur optimal. On utilisera la statistique de test donnée par le corollaire suivant.

Corollaire 1.4.1. Sous les hypothèses du théorème précédent et sous l'hypothèse $H_0 \ : \ r_0 = 2$, on a

$$\widehat{\zeta}_n = \frac{n}{\widehat{\tau}_2}(\widehat{r}_n - 2)^2 \xrightarrow{\mathcal{L}} \chi^2(1).$$

On peut maintenant mettre en place une stratégie d'estimation du paramètre θ_0. On commence par calculer $\hat{\theta}_{n,2}$. À partir de cette quantité, on estime $\hat{r}_n$ et on teste l'hypothèse $H_0 \ : \ r_0 = 2$. Si on ne peut pas rejeter cette hypothèse, alors on conserve les estimations du QML gaussiens, sinon on estime $\hat{\theta}_{n,\hat{r}_n}$ qui devrait asymptotiquement fournir de meilleures estimations.

Des simulations numériques sont ensuite présentées démontrant que l'estimateur du QMV non gaussien en deux étapes fournit, dans certains cas de meilleurs estimations que le QMV gaussien. En particulier, on peut citer le cas où le processus $(\eta_t)_t$ est distribué selon une loi de Student de faible degré (et est donc fortement leptokurtique) où l'estimation par QMV non gaussien donne des résultats spectaculairement meilleurs que le QMV gaussien.

1.5 Résultats du chapitre 3

Dans ce chapitre, on étudie les propriétés asymptotiques de l'estimateur du maximum de vraisemblance d'un modèle conditionnellement hétéroscédastiques lorsque le processus des innovations est distribué selon une loi alpha stable. On généralise ensuite les résultats obtenus au cas où les innovations ne sont plus supposées distribuées selon une loi alpha stable mais convergeant en distribution vers une telle loi. On prouve ainsi la robustesse de l'estimateur du maximum de vraisemblance à ce type d'erreur de spécification.

L'étude des modèles conditionnellement hétéroscédastiques lorsque le processus des innovations est leptokurtique est extrêmement important. En effet, les performances du QMV gaussien peuvent être altérées par la présence d'innovations à queues épaisses. Les séries financières sur lesquelles on applique ces modèles présentent le plus souvent une forte leptokurticité, indiquant ainsi que l'on peut avoir besoin d'utiliser une autre méthode d'estimation. Certains chercheurs ont étudié l'estimation de modèles GARCH en présence d'innovations non gaussiennes. On peut citer Berkes and Horváth (2004) pour une approche générale. Dans le cas où le processus $(\eta_t)_t$ n'admet pas de moment d'ordre 4 mais admet une variance finie, Hall and Yao (2003) montrent que l'estimateur du QMV est consistent mais ils obtiennent une loi asymptotique non gaussienne et une vitesse de convergence plus lente que $\sqrt{n}$ et qui doit être approximée par des méthodes utilisant le bootstrap. Les modèles avec innovations stables ont déjà été étudié dans la littérature, ainsi le modèle GARCH avec des innovations stables a été étudié empiriquement par Liu and Brorsen (1995). Les propriétés asymptotiques d'un estimateur du MV de modèles ARMA avec innovations stables ont été établies théoriquement par Andrews et al. (2009). On peut remarquer que si des résultats existent au sujet de l'estimateur du maximum de vraisemblance de modèles conditionnellement hétéroscédastiques (voir par exemple Straumann (2005)), le cas des lois alpha stables, plus compliqué n'a pas été traité. Il convient dans un premier temps de donner quelques propriétés des lois alpha stables qui pourront motiver leur emploi.

1.5.1 Propriétés des lois alpha stables

Définition 1.5.1. Une loi de probabilité est dite alpha stable, si pour un échantillon de taille n, $(Z_t)_{t=1\cdots n}$ de cette distribution, il existe des quantités $a_n > 0$ et b_n telles que

$$\frac{Z_1 + \cdots + Z_n}{a_n} - b_n \overset{\mathcal{L}}{=} Z_1.$$

La principale difficulté avec les lois alpha stables est qu'il n'existe pas d'écriture analytique de la densité de probabilité. On peut seulement écrire la forme de la fonction caractéristique. La loi alpha stable est décrite par quatre paramètres :

- Le paramètre α est le paramètre de queue, il est compris entre 0 et 2. Plus il est proche de 0, plus la loi est leptokurtique.
- Le paramètre $\beta \in (-1, 1)$ est un paramètre d'asymétrie. S'il est égal à 1 ou -1 et si $\alpha < 1$ alors le support de la distribution est borné d'un côté de la droite des réels.
- Le paramètre $\mu \in \mathbb{R}$ est un paramètre d'emplacement Il ne peut cependant pas être identifié à la moyenne car certaines lois alpha stables ne possèdent pas de moyenne.
- Le paramètre $\gamma \in \mathbb{R}_+$ est un paramètre d'échelle.

Avec ces paramètres et en utilisant la paramétrisation de Zolotarev (1986), on peut écrire la fonction caractéristique. Une variable X est dite alpha stable de paramètre $\psi = (\alpha, \beta, \mu, \gamma)$ si sa fonction caractéristique s'écrit

$$E\left[\exp itX\right] = \begin{cases} \exp\left\{-|\gamma t|^\alpha + \gamma^\alpha i\beta \tan\left(\frac{\alpha\pi}{2}\right) t\left(|t|^{\alpha-1} - 1\right)\right\} + i\mu t & \text{si} \quad \alpha \neq 1 \\ \exp\left\{-|\gamma t| - \gamma i\beta t \frac{2}{\pi} \log|\gamma t|\right\} + i\mu t & \text{si} \quad \alpha = 1. \end{cases}$$

À partir de cette fonction caractéristique et en utilisant la transformée inverse de Fourier ou des résultats d'intégrations, on pourra numériquement évaluer la densité $f(., \psi)$ d'une loi alpha stable. En utilisant toujours cette fonction caractéristique, on peut également donner le comportement de la densité f et de ses dérivées partielles dans les queues de distribution On obtient ainsi quand $|x| \to +\infty$

$$f(x, \psi) \sim K x^{-\alpha-1},$$

où K est une constante pouvant prendre différentes valeurs en fonction du signe de la limite de x. Ce résultat nous permet de remarquer qu'une variable aléatoire alpha stable de paramètres $\psi = (\alpha, \beta, \mu, \gamma)'$ n'admet que des moments d'ordre $s < \alpha$. Ces lois sont donc extrêmement leptokurtiques.

La dernière propriété des lois alpha stables que nous exposons est celle qui motive leur utilisation. En effet, ce sont les seules lois à posséder un domaine d'attraction, c'est-à-dire que pour un échantillon $(X_t)_{t=1,\cdots,n}$, s'il existe des quantités $a_n > 0$ et b_n et une variable aléatoire Y telles que

$$\frac{1}{a_n} \sum_{t=1}^{n} X_t - b_n \overset{\mathcal{L}}{\to} Y, \text{ quand } n \to +\infty,$$

alors Y est nécessairement de loi alpha stable. Grâce à cette propriété, on vérifie aisément que la loi normale est une loi alpha stable correspondant au cas particulier $\alpha = 2$. En effet, avec le théorème de la limite centrale, toutes les lois possédant une variance finie sont dans le domaine d'attraction de la loi gaussienne. Ce théorème peut être généralisé au cas où la variable sommée admet une variance infinie. On trouve en effet dans Gnedenko et al. (1968) le théorème suivant.

Théorème 1.5.1. Pour un processus (X_t) i.i.d. et vérifiant

$$\mathbb{P}\left[X_t > x\right] \sim K_1 x^{-\alpha} \text{ quand } x \to +\infty, \tag{1.5.1}$$
$$\mathbb{P}\left[X_t < x\right] \sim K_2 |x|^{-\alpha} \text{ quand } x \to -\infty, \tag{1.5.2}$$

avec $\alpha \in (0, 2)$, $K_1 > 0$ et $K_2 > 0$, alors, en posant

$$\beta = \frac{K_1 - K_2}{K_1 + K_2}, \ a = \left\{-\alpha M(\alpha)(K_1 + K_2)\cos\frac{\alpha\pi}{2}\right\}^{1/\alpha},$$
$$M(\alpha) = \begin{cases} -\dfrac{\Gamma(1-\alpha)}{\alpha}, \text{ quand } \alpha < 1 \\ \dfrac{\Gamma(2-\alpha)}{\alpha(\alpha-1)}, \text{ quand } \alpha > 1, \end{cases}$$

on a

$$\frac{1}{an^{1/\alpha}} \sum_{t=1}^{n} (X_t - m) \xrightarrow{\mathcal{L}} Z,$$

avec $m = EX_1$ quand $\alpha > 1$, $m = 0$ quand $\alpha < 1$ et $Z \sim S(\alpha, \beta, \beta \tan\frac{\alpha\pi}{2}, 1)$.

On trouve également chez Basu and Maejima (1980) une extension de ce théorème, avec une hypothèse en plus, ils obtiennent également la convergence en densité de la somme vers une loi stable, si f_n est la densité de $\frac{1}{an^{1/\alpha}} \sum_{t=1}^{n} (X_t - m)$, ils obtiennent alors pour $0 \leq \delta \leq \alpha$,

$$\sup_{x \in \mathbb{R}} (1 + |x|)^{\delta} |f_n(x) - f(x, \psi)| \to 0, \quad \text{quand } n \to +\infty.$$

Ces résultats nous permettront d'étudier le cas où le processus des innovations du modèle GARCH s'écrit comme une somme de variables indépendantes vérifiant (1.5.1) et (1.5.2).

1.5.2 Maximum de vraisemblance d'un modèle GARCH avec innovations alpha stables

On étudie le modèle (1.2.2) en spécifiant la distribution du processus $(\eta_t)_t$ comme une loi alpha stable de paramètre ψ_0. La paramètre du modèle est alors $\tau_0 = (\theta_0', \psi_0')'$. Pour des raisons d'identifiabilité, le paramètre γ_0 doit être fixé à 1 et est ainsi omis des estimations, on pose donc $\psi_0 = (\alpha_0, \beta_0, \mu_0)'$. On définit l'estimateur du maximum de vraisemblance. Si Γ représente l'espace des paramètres, alors pour $\tau = (\theta', \psi')' \in \Gamma$, on définit le critère à minimiser :

$$\tilde{I}_n(\tau) = \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\tau) \quad \text{où} \quad \tilde{l}_t(\tau) = \frac{1}{2} \log \tilde{\sigma}_t^2(\theta) - \log f\left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}, \psi\right).$$

Puis on définit τ_n l'estimateur du maximum de vraisemblance du modèle (1.2.2) dans le cas "alpha stable" :

$$\tau_n = \underset{\tau \in \Gamma}{\operatorname{argmin}} \ \tilde{I}_n(\tau).$$

Pour pouvoir étudier τ_n, on doit se placer dans le cas stationnaire, on fait donc l'hypothèse

A0 (ϵ_t) est une solution causale, strictement stationnaire et ergodique du modèle (1.2.2).

Cette hypothèse est bien plus contraignante sur l'espace des paramètres que cette même hypothèse faite dans le cas où le processus des innovations est gaussien. En effet si la persistance du processus σ_t est trop forte et si le paramètre α_0 est trop faible alors $(\epsilon_t)_t$ sera explosif.

Lorsque l'on se place dans le cas du modèle conditionnellement hétéroscédastique général, on doit énoncer de nombreuses hypothèses pour obtenir la consistance et la normalité asymptotique de τ_n (il s'agit des hypothèses **A1-A8** du chapitre 3), ces hypothèses sont facilement vérifiées dans le cas particulier du modèle GARCH. La seule hypothèse restante est une hypothèse sur l'espace des paramètres Γ dont il faut supposer la compacité. On obtient ainsi le théorème suivant.

Théorème 1.5.2. Sous les hypothèses idoines, τ_n est un estimateur consistant de τ_0

$$\tau_n \underset{n\to+\infty}{\longrightarrow} \tau_0 \quad \text{a.s.} \tag{1.5.3}$$

On obtient également la normalité asymptotique de cet estimateur.

$$\sqrt{n}(\tau_n - \tau_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, J^{-1}\right),$$

où $J = E\left[\frac{\partial^2 l_t(\tau_0)}{\partial\tau\partial\tau'}\right] = E\left[\frac{\partial l_t}{\partial\tau}(\tau_0)\frac{\partial l_t}{\partial\tau'}(\tau_0)\right]$, avec $l_t(\tau) = \frac{1}{2}\log\sigma_t^2(\theta) - \log f\left(\frac{\epsilon_t}{\sigma_t(\theta)}, \psi\right)$.

1.5.3 Cas où le processus des innovations converge en distribution vers une loi stable

Dans cette partie, on cherche un équivalent à la propriété de robustesse du QML gaussien. En effet, lorsque le processus des innovations est dans le domaine d'attraction d'une loi gaussienne, c'est-à-dire possède une variance finie, alors le QMLE gaussien est consistant. Sachant que la loi gaussienne est un cas particulier de la famille des lois alpha stables, on voudrait généraliser cette propriété au cas où l'innovation est dans le domaine d'attraction d'une loi alpha stable avec $\alpha < 2$. On considère donc ici un modèle similaire au modèle (1.2.2) avec un processus des innovations qui est cette fois indexé par n et qui converge en distribution vers une loi alpha stable de paramètre ψ_0. On peut écrire formellement ce modèle de la façon suivante.

$$\begin{cases} \epsilon_{nt} = \sigma_{nt}\eta_{nt} \\ \sigma_{nt}^2 = \omega_0 + \sum_{i=1}^{q} a_{0i}\epsilon_{nt-i}^2 + \sum_{j=1}^{p} b_{0j}\sigma_{nt-j}^2, \quad \forall t \in \mathbb{Z},\ \forall n \in \mathbb{N}, \end{cases} \tag{1.5.4}$$

Le processus $(\eta_{nt})_t$ converge en distribution vers une loi alpha stable de paramètre ψ_0 quand $n \to +\infty$, cela se produit si par exemple η_{nt} s'écrit

$$\eta_{nt} = \frac{1}{k_n}\sum_{i=1}^{k_n} \nu_{it},$$

où $(k_n)_n$ est une suite croissante avec n et où ν_{it} vérifie les hypothèses du théorème 1.5.1.

On utilise le même estimateur que dans la partie précédente, on maximise la vraisemblance calculée sous l'hypothèse que le processus des innovations suit une loi stable. Dans ce cas précis, cette hypothèse est fausse puisque $(\eta_{nt})_t$ n'est pas alpha stable. L'estimateur défini est donc un estimateur du pseudo maximum de vraisemblance. On définit maintenant γ comme le coefficient de Lyapunov associé au modèle (1.2.2) dans le cas où $(\eta_t)_t$ suit une loi stable. On définit également pour tout $n \in \mathbb{N}$, γ_n le coefficient de Lyapunov associé au modèle (1.5.4). On énonce maintenant les hypothèses qui seront requises pour dériver les propriétés asymptotiques de l'estimateur du pseudo maximum de vraisemblance.

B1 $\tau_0 \in \Gamma$ et Γ est un espace compact.

B2 $\gamma < 0$ et $\forall \theta \in \Theta,\ \sum_{j=1}^{p} b_j < 1$.

B3 Il existe $\delta > 1$ vérifiant, pour tout $n \in \mathbb{N}$, $E|\eta_{nt}|^{\delta} < +\infty$ et $\sup_{x\in\mathbb{R}} (1+|x|)^{\delta}|f_n(x) - f(x,\psi_0)| \to 0$.

B4 Si $p > 0$, $\mathcal{A}_{\theta_0}(z)$ et $\mathcal{B}_{\theta_0}(z)$ ne possèdent pas de racine commune, $\mathcal{A}_{\theta_0}(1) \neq 0$ et $a_{0q} + b_{0p} \neq 0$.

B5 On a $\sup_{n\in\mathbb{N}} \alpha_{\epsilon_n}(h) \leq K\rho^h$, où $\alpha_{\epsilon_n}(h)$ pour $h \in \mathbb{N}$ est défini comme le coefficient de α-mélange du processus (ϵ_{nt}).

B6 $\tau_0 \in \overset{\circ}{\Gamma}$, où $\overset{\circ}{\Gamma}$ représente l'intérieur de Γ.

On note que l'on n'a pas besoin de supposer que les coefficients de Lyapunov associés au modèle pour un n donné sont négatifs. En effet l'hypothèse **B2** implique que pour un $N \in \mathbb{N}$, on a

$$n \geq N \Rightarrow \gamma_n < 0.$$

On peut se reporter au lemme 7.4 et à sa preuve pour plus de détails..

Sous les hypothèses énoncée précédemment, on obtient la convergence et la normalité asymptotique de l'estimateur de pseudo maximum de vraisemblance. Il est intéressant de noter que la matrice de variance covariance asymptotique est la même que dans le cas bien spécifié. Il n'y a pas de coût asymptotique à ne pas spécifier la distribution du processus des innovations mais à seulement spécifier la limite en loi de ce processus. Le théorème s'énonce de la façon suivante

Théorème 1.5.3. Sous les hypothèses **B1-B5**, l'estimateur τ_n est consistant,

$$\tau_n \underset{n\to+\infty}{\longrightarrow} \tau_0,\ a.s.$$

Si on a ausi l'hypothèse **B6**, alors,

$$\sqrt{n}(\tau_n - \tau_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, J^{-1}\right),$$

avec $J = E\left[\frac{\partial^2 l_t(\tau_0)}{\partial\tau\partial\tau'}\right] = E\left[\frac{\partial l_t}{\partial\tau}(\tau_0)\frac{\partial l_t}{\partial\tau'}(\tau_0)\right]$.

On donne maintenant quelques idées de la preuve de ce théorème. Cette preuve suit le schéma de la preuve du théorème 1.5.2 à la différence que toutes les quantités utilisées ne dépendent plus du processus $(\eta_t)_t$ mais de $(\eta_{nt})_t$. Cette dépendance en n sera la principale difficulté à traiter. Il faudra prouver les convergences dans L^1 des quantités dépendant de n vers leurs équivalents limites. On détaille ici l'exemple du Lemme 7.7 du chapitre 3. Dans ce lemme, on prouve

$$E\left[\inf_{\tau\in\Gamma} l_{nt}(\tau)\right] \underset{n\to+\infty}{\longrightarrow} E\left[\inf_{\tau\in\Gamma} l_t(\tau)\right], \tag{1.5.5}$$

avec $l_t(\tau) = \frac{1}{2}\log\sigma_t^2(\theta) - \log f\left(\frac{\epsilon_t}{\sigma_t(\theta)}, \psi\right)$ et $l_{nt}(\tau) = \frac{1}{2}\log\sigma_{nt}^2(\theta) - \log f\left(\frac{\epsilon_{nt}}{\sigma_{nt}(\theta)}, \psi\right)$. La quantité $l_{nt}(\tau)$ est construite à partir du processus des innovations $(\eta_{nt})_t$ et est définie pour tout $n \in \mathbb{N}$. La quantité $l_t(\tau)$ est quant à elle construite à partir du processus limite $(\eta_t)_t$. À partir principalement de l'hypothèse **B3** on veut montrer la convergence L^1 de l'équation (1.5.5). La principale difficulté provient du fait que la quantité σ_{nt} peut s'écrire comme une fonction d'un nombre infini de $\eta_{nt'}$. En conséquence, on ne peut pas appliquer le théorème de convergence dominée en utilisant l'hypothèse **B3**. On doit donc introduire une version tronquée de σ^2_{nt} que l'on note $\sigma^{2(m)}_{nt}$. Pour simplifier la situation et donner l'idée de cette preuve, on définit

$$\underline{z}_{nt} = \left(\epsilon^2_{nt}, \cdots, \epsilon_{nt-q+1}, \sigma^2_{nt}, \cdots, \sigma_{nt-p+1}\right)' \in \mathbb{R}^{p+q}.$$

On montre dans le chapitre 3 que cette quantité peut s'écrire de la façon suivante

$$\underline{z}_{nt} = \sum_{k=0}^{+\infty} A_{nt} \cdots A_{nt-k+1} \underline{b}_{nt-k},$$

où pour tout $t \in \mathbb{R}$, la matrice A_{nt} n'est fonction que de η_{nt}. On définit maintenant une version tronquée de $\underline{z}_{nt}$, en posant, pour $m \in \mathbb{N}$

$$\underline{z}^{(m)}_{nt} = \sum_{k=0}^{m} A_{nt} \cdots A_{nt-k+1} \underline{b}_{nt-k}.$$

Ainsi définie, $\underline{z}^{(m)}_{nt}$ peut s'écrire comme une fonction de $(\eta_{nt})_{t\in\{1,\cdots,m\}}$. Le critère l_{nt} étant fonction de σ^2_{nt} qui est une composante de $\underline{z}_{nt}$, on peut également définir une version tronquée de l_{nt} que l'on appellera $l^{(m)}_{nt}$. Pour obtenir (1.5.5), on démontrera successivement les trois résultat suivants.

(i) $\sup_{n\in\mathbb{N}} E \left| \inf_{\tau\in V} l_{nt}(\tau) - \inf_{\tau\in V} l^{(m)}_{nt}(\tau) \right| < K\rho^m$.

(ii) $E \left| \inf_{\tau\in V} l_t(\tau) - \inf_{\tau\in V} l^{(m)}_t(\tau) \right| < K\rho^m$.

(iii) Pour tout $m > 0$, $E \inf_{\tau\in V} l^{(m)}_{nt}(\tau) \to E \inf_{\tau\in V} l^{(m)}_t(\tau)$, quand $n \to +\infty$,

où V est défini comme un sous espace de l'espace des paramètres Γ.

Les deux premiers résultats s'obtiennent en prouvant le lemme suivant (Lemme 7.6 du chapitre 3)

$$\sup_{n\in\mathbb{N}} \sup_{\theta\in\Theta} E \left| \sigma^{2s}_{nt}(\theta) - \sigma^{2(m)s}_{nt}(\theta) \right| < K\rho^m,$$

où K et ρ sont des constantes vérifiant $K > 0$ et $0 < \rho < 1$.

Le point (iii) s'obtient en appliquant le théorème de convergence dominée. En effet, si ce théorème ne peut s'appliquer à $l_{nt}(\tau)$, la version tronquée $l^{(m)}_{nt}$ ne dépend que d'un nombre fini de $(\eta_{nt})_t$, ainsi en utilisant en particulier l'hypothèse **B3**, on pourra prouver le résultat. Des résultats intermédiaires (i), (ii) et (iii) on pourra ensuite obtenir (1.5.5).

1.5.4 Simulations et application à des données réelles.

L'estimateur du pseudo maximum de vraisemblance stable est ensuite testé numériquement. On simule des réalisations d'un modèle GARCH avec un processus des innovations pouvant s'écrire de la façon suivante, pour $K > 0$

$$\eta_t^{(K)} = \frac{1}{K^{1/\alpha}} \sum_{k=1}^{K} \nu_{k,t}, \quad \text{avec} \quad (\nu_{k,t})_k \overset{iid}{\sim} t_\alpha.$$

Ainsi défini, le processus $\left(\eta_t^{(K)}\right)_t$ converge en distribution lorsque $K \to +\infty$ vers une loi stable dont les paramètres peuvent être déterminé, voit Théorème 1.5.1. Le cas $K = +\infty$ correspond au cas où le processus des innovations est distribué selon un loi stable. Pour différentes valeurs de K, nous étudions le comportement de l'estimateur du pseudo maximum de vraisemblance stable (le cas $K = +\infty$ correspond alors à une estimation par exact maximum de vraisemblance). Si l'estimation des paramètres du modèle GARCH semble devenir plus efficace à mesure que la valeur de K augmente, nous trouvons que même pour $K = 500$, l'estimation des paramètres dans le cas mal spécifié est comparable à l'estimation dans le cadre du maximum de vraisemblance.

Nous étudions ensuite l'estimation d'un tel modèle sur plusieurs séries financières. Les paramètres α estimés sont plus petits que 2, ce qui justifie l'utilisation de la loi stable. Nous comparons également les *Value-at-Risk* (mesure de risque que nous définissons dans le chapitre 3) obtenues par cette modélisation avec les Value-at-Risk obtenues par l'estimation d'un modèle GARCH avec innovation gaussienne. Nous montrons que si ces deux méthodes semblent donne des résultats similaires pour un niveau de 5%, la Value-at-Risk stable est bien plus performante pour des niveaux inférieurs. La modélisation stable semble donc mieux rendre compte du comportement dans les queues de distribution du processus.

1.6 Résultats du chapitre 4

Dans ce chapitre, on étudie les résultats de l'estimation d'un modèle mal spécifié. Ici la mauvaise spécification ne se fait pas simplement sur la densité du processus des innovations, mais se fait sur le modèle en lui même. Le vrai modèle suivi par le processus $(\epsilon_t)_t$ sera dans ce chapitre un modèle GARCH à changement de régimes Markoviens (MS-GARCH).

1.6.1 Le modèle MS-GARCH

Ce modèle est semblable au modèle (1.2.2), à la différence que les coefficients sont ici dépendants du temps. Leur dynamique est régie par une chaîne de Markov Δ_t. Ce modèle s'écrit plus formellement

$$\begin{cases} \epsilon_t = \sqrt{h_t} u_t \\ h_t = w(\Delta_t) + \sum_{i=1}^{q} a_i(\Delta_t)\epsilon_{t-i}^2 + \sum_{j=1}^{p} b_j(\Delta_t) h_{t-j}, \end{cases} \tag{1.6.1}$$

La chaîne $(\Delta_t)_t$ est supposée indépendante des innovations $(u_t)_t$. Elle est également supposée ergodique. On définit les quantités suivantes,

$$A_t = \begin{pmatrix} a_1(\Delta_t)u_t^2 + b_1(\Delta_t) & & \cdots & a_r(\Delta_t)u_{t-r}^2 + b_r(\Delta_t) \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & \cdots \; 1 & 0 \end{pmatrix}, \tag{1.6.2}$$

ainsi que

$$\underline{z}_t = \begin{pmatrix} h_t \\ \vdots \\ h_{t-p+1} \end{pmatrix}, \ \underline{b}_t = \begin{pmatrix} w(\Delta_t) \\ \vdots \\ 0 \end{pmatrix} \ \text{et } B_t = \begin{pmatrix} b_1(\Delta_t) & & \cdots & & b_q(\Delta_t) \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & & 1 & 0 \end{pmatrix}.$$

Avec ces notations, on peut écrire une version vectorielle du modèle (1.6.1)

$$\underline{z}_t = \underline{b}_t + A_t \underline{z}_{t-1}. \tag{1.6.3}$$

Les probabilités stationnaires de la chaîne $(\Delta_t)_t$ sont notées $\pi(k) = \mathbb{P}[\Delta_1 = k]$. Les probabilités de transition sont quant à elles notées $p(k,l) = \mathbb{P}[\Delta_t = l | \Delta_{t-1} = k]$ pour $k, l \in \{1, \cdots, d\}$. Enfin, pour une fonction $f : \{1, \cdots, d\} \mapsto \mathcal{M}_{n \times n'}(\mathbb{R})$, on définit

$$\mathbb{P}_f = \begin{pmatrix} p(1,1)f(1) & \cdots & p(d,1)f(1) \\ \vdots & & \vdots \\ p(d,1)f(d) & \cdots & p(d,d)f(d) \end{pmatrix} \quad \text{et} \quad \Pi_f = \begin{pmatrix} \pi(1)f(1) \\ \vdots \\ \pi(d)f(d) \end{pmatrix}.$$

On énonce dans ce chapitre une nouvelle condition de stationnnarité différente de ce qui existe dans la littérature traitant des modèles MS-GARCH. En effet, on trouve habituellement (voir Francq and Zakoïan (2005), Liu (2006) et Mittnik et al. (2002)) des conditions de stationnarité qui impliquent l'existence d'un moment d'ordre 2 pour le processus et qui impliquent donc la stationnarité au second-ordre. On trouve également des conditions de stationnarité stricte utilisant le coefficient de Lyapunov mais on ne peut pas prouver comme dans le cas du modèle GARCH que cela implique l'existence d'un $s > 0$ tel que $E|\epsilon_t^{2s}| < +\infty$. Dans ce papier, on estime un modèle GARCH mal spécifié, le processus qui génère les données n'est pas un GARCH mais un MS-GARCH. On veut donc que les modèles soient comparables et partagent des propriétés communes. Usuellement on ne suppose que l'existence d'un moment d'ordre $2s$ (avec $s > 0$) pour le processus $(\epsilon_t)_t$, on veut donc ici une condition de stationnarité impliquant seulement l'existence d'un tel moment. On fait donc ici l'hypothèse suivante. On définit auparavant $\rho\left(\mathbb{P}_{\tilde{A}^s}\right)$ comme le rayon spectral de la matrice $\tilde{A}^s$ définie par $\tilde{A}^s(k) = E\left[A_t^{(s)} | \ \Delta_t = k\right]$, où pour une matrice quelconque M, la matrice $M^{(s)}$ est définie par $M^{(s)} = (M(i,j)^s)_{(i,j)}$.

A0 Il existe $s \in (0,1)$ tel que $E|u_t|^{2s} < +\infty$ et $\rho\left(\mathbb{P}_{\tilde{A}^s}\right) < 1$, ou pour une matrice générique M, $\rho(M)$ est le rayon spectral de cette matrice.

Cette condition est suffisante pour obtenir l'existence d'une solution stationnaire de (1.6.1) et pour obtenir un moment d'ordre $2s$ pour $(\epsilon_t)_t$. On peut également noter que cette condition implique la condition plus classique de stationnarité, $\gamma < 0$ où γ est le coefficient de Lyapunov associé au modèle (1.6.1). Dans ce chapitre, la stationnarité est particulièrement importante. En effet nous étudierons séparément l'estimation d'un modèle GARCH$(1,1)$ sur un modèle MS-GARCH dans le cas stationnaire ainsi que dans le cas explosif et donc non stationnaire. On peut noter que dans le cas explosif, il n'est pas possible d'estimer le paramètre ω, aussi nous ne donnerons des résultats asymptotiques que pour les estimateurs des quantités α_0 et β_0.

1.6.2 Estimation d'un GARCH(1,1) sur un MS-GARCH stationnaire

Dans cette section, le DGP est un processus MS-GARCH(p,q) et on se place dans le cas où l'on estime un mauvais modèle. Le modèle le plus utilisé pour des séries conditionnellement hétéroscédastiques est très largement le modèle GARCH$(1,1)$. C'est donc ce modèle que l'on estime sur le DGP. On utilise l'estimateur du quasi maximum de vraisemblance et on définit comme dans les sections précédentes le critère à minimiser,

$$\tilde{I}_n(\theta) = \sum_{t=1}^{n} \tilde{l}_t(\theta), \quad \tilde{l}_t(\theta) = \log \tilde{\sigma}_t^2(\theta) + \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\theta)},$$

où les $\tilde{\sigma}_t$ sont définis récursivement en utilisant des valeurs initiales. On aura besoin des hypothèses suivantes,

A1 Pour tout $\theta \in \Theta,\ \beta < 1$.

A2 Θ est un espace compact et $\rho(\mathbb{Q}_B) < 1$.

A3 $\Theta^* \neq \emptyset$, où $\Theta^* = \{\theta^* = (\omega^*, \alpha^*, \beta^*) \in \Theta,\ \rho(\mathbb{Q}_B) < \beta^* < 1\}$.

A4 Il existe un unique $\theta_0 = (w_0, \alpha_0, \beta_0)' \in \Theta$ tel que $\theta_0 = \underset{\theta\in\Theta}{\operatorname{argmin}}\ El_t(\theta)$ et $\theta_0 \in \overset{\circ}{\Theta}$

A5 Il existe $\eta_1 > 0$ tel que $E\left|l_t(\theta_0)\right|^{1+\eta_1} < +\infty$.

A6 Il existe $\eta_2 > 0$ tel que $Eu_t^{4+\eta_2} < +\infty$.

A7 La matrice $A(\theta_0)$ est inversible, en définissant $A(\theta_0) = E\left[\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right]$.

Avec ces hypothèses, on obtient la consistance de l'estimateur du modèle mal spécifié vers la pseudo vraie valeur θ_0

Théorème 1.6.1. Sous les hypothèses **A0-A4**, on a

$$\theta_n \underset{n\to+\infty}{\longrightarrow} \theta_0,\ a.s.$$

Ce résultat découle principalement de l'hypothèse **A4** en suivant un schéma classique de preuve de consistance. On obtient également la loi asymptotique de cet estimateur.

Théorème 1.6.2. Sous les hypothèses **A0-A6**, on obtient

$$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-1}\right),$$

où $A(\theta_0)$ et $B(\theta_0)$ sont des matrices définies positives définies par

$$B(\theta) = E\left[\frac{\partial l_t(\theta)}{\partial\theta}\frac{\partial l_t(\theta)}{\partial\theta'}\right], \quad A(\theta) = E\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right].$$

Dans les preuves de ce chapitre, on définit, pour $\theta \in \Theta$ la quantité $\sigma_t(\theta)$ comme un équivalent stationnaire du processus $\tilde{\sigma}_t(\theta)$. $\sigma_t(\theta)$ correspond au cas où la récurrence serait initialisée en $t = -\infty$.

Le théorème 1.6.2 est plus délicat à obtenir car on n'a pas d'indépendance comme dans le cas bien spécifié. Habituellement, lorsque les quantités impliquées dans les preuves sont évaluées en $\theta = \theta_0$, on a des termes de la forme suivante

$$\left(1 - \frac{\epsilon_t^2}{\sigma_t^2(\theta_0)}\right) \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2}{\partial \theta}(\theta_0) = (1 - \eta_t^2) \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2}{\partial \theta}(\theta_0).$$

Dans le cas bien spécifié on a indépendance entre les quantité $(1 - \eta_t^2)$ et $\frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2}{\partial \theta}(\theta_0)$. Dans le cas mal spécifié, on n'a pas de propriété d'indépendance entre $\frac{\epsilon_t^2}{\sigma_t^2(\theta_0)}$ et les quantités faisant apparaitre $\sigma_t^2(\theta_0)$ ou ses dérivées. Les preuves en sont donc significativement compliquées.

1.6.3 Estimation d'un GARCH(1,1) sur un MS-GARCH non stationnaire.

On se place maintenant dans le cas où l'hypothèse **A0** n'est pas vérifiée. On suppose que le coefficient de Lyapunov γ associé au modèle est strictement plus grand que 0. On montre alors que cette hypothèse implique que le modèle est explosif, plus précisément, on prouve que pour tout $\rho > e^{-\gamma}$, on a quand $t \to +\infty$

$$\rho^t h_t \to +\infty, \text{ a.s.}$$

Cette propriété est extrêmement importante et sera utilisée à chaque étape des preuves de cette section.

On estime donc un modèle GARCH$(1,1)$ lorsque le DGP est un processus MS-GARCH explosif. On peut montrer que l'estimateur du QMV s'obtient en minimisant la quantité suivante.

$$I_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left\{ u_t^2 \left(\frac{h_t}{\sigma_t^2(\theta)} - 1 \right) + \log \frac{\sigma_t^2(\theta)}{h_t} \right\}.$$

On peut ici noter un changement de notation, en effet les quantités $\sigma_t^2(\theta)$ étant explosives quand $t \to +\infty$ il n'est pas possible de leur écrire un équivalent stationnaire. On n'utilise donc pas ici la notation $\tilde{\sigma}_t$. On peut cependant définir une alternative stationnaire à la quantité $\frac{\sigma_t^2(\theta)}{h_t}$, on pose

$$v_t(\alpha, \beta) = \sum_{j=1}^{+\infty} \alpha \beta^{j-1} u_{t-j}^2 \prod_{k=1}^{j-1} \frac{1}{a_{t-k}},$$

et on prouve le résultat suivant (Lemme 5.8 du chapitre 4). Si on définit le sous espace $\Theta_\gamma = \{\theta \in \Theta,\ \beta < e^{\gamma}\}$, alors sous les hypothèses idoines et pour tout $\theta \in \Theta_\gamma$, $v_t(\alpha, \beta)$ est stationnaire et ergodique. De plus, pour tout compact $\Theta_\gamma^* \subset \Theta_\gamma$, on a

$$\sup_{\theta \in \Theta_\gamma^*} \left| \frac{\sigma_t^2(\theta)}{h_t} - v_t(\alpha, \beta) \right| \to 0 \ a.s. \text{ quand } t \to +\infty.$$

Dans le cas $\theta \notin \Theta_\gamma$, on obtient

$$\frac{\sigma_t^2(\theta)}{h_t} \to +\infty, \text{ p.s. quand } t \to +\infty.$$

On définit les hypothèses suivantes

B0 $\gamma > 0$.

B1 Il existe un unique couple (α_0, β_0) tel que

$$(\alpha_0, \beta_0) = \underset{(\alpha,\beta)}{\text{argmin}}\ E\left[u_t^2\left(\frac{1}{v_t(\alpha,\beta)} - 1\right) + \log v_t(\alpha,\beta)\right].$$

B2 L'espace des paramètres Θ est compact.

B3 Le processus des innovations (u_t) vérifie $\mathbb{P}[u_t = 0] = 0$.

Sous ces hypothèses, on obtient un résultat de convergence pour les quantités α_n et β_n. Le modèle étant explosif, il n'est pas possible d'obtenir un résultat de convergence pour ω_n.

Théorème 1.6.3. Sous les hypothèses **B0-B3**, on obtient

$$\alpha_n \to \alpha_0,\ \text{ et } \beta_n \to \beta_0,\ \text{ a.s. quand } n \to +\infty.$$

Comme on a définit un "équivalent" stationnaire de $\frac{\sigma_t^2(\theta)}{h_t}$, on peut définir des équivalents des quantités $\frac{1}{h_t}\frac{\partial\sigma_t^2}{\partial\theta_i}(\theta)$. En utilisant ces quantités et avec les hypothèses suivantes, on pourra trouver la loi asymptotique du couple (α_n, β_n). On définit auparavant les quantités suivantes. Pour $j \in \mathbb{N}$ et $k \in \{1, \cdots, d\}$, on donne la fonction $\tilde{a}^{(j)}:\ k \mapsto E\left[\frac{1}{\left(a(k)u_1^2 + b(k)\right)^j}\right]$ et avec cette notation on définit $\Theta^{(j)} = \left\{\theta \in \Theta_\gamma,\ \beta^j \rho(\mathbb{P}_{\tilde{a}^{(j)}}) < 1\right\}$. On pose également pour $i, j \in \{2, 3\}$,

$$D_t^{\theta_i,\theta_j} = \frac{\partial v_t}{\partial\theta_i\partial\theta_j}(\alpha_0,\beta_0)\left(1 - \frac{u_t^2}{v_t(\alpha_0,\beta_0)}\right) + \frac{\partial v_t}{\partial\theta_i}(\alpha_0,\beta_0)\frac{\partial v_t}{\partial\theta_j}(\alpha_0,\beta_0)\left(2\frac{u_t^2}{v_t(\alpha_0,\beta_0)} - 1\right),$$

ainsi que $C_{\alpha_0,\beta_0} = E\begin{pmatrix} D_1^{\alpha,\alpha} & D_1^{\alpha,\beta} \\ D_1^{\alpha,\beta} & D_1^{\beta,\beta} \end{pmatrix}$.

B4 Pour tout $\omega_0 > 0$, on a $\theta_0 = (\omega_0, \alpha_0, \beta_0) \in \Theta^{(1)} \cap \Theta^{(2)}$.

B5 Il existe $\tilde{\theta} = (\tilde{\omega}, \tilde{\alpha}, \tilde{\beta})'$ tel que, $\tilde{\theta} \in \bigcap_{j=1}^{+\infty} \Theta^{(j)}$ et $\tilde{\beta} > 1$.

B6 Il existe $\eta_3 > 0$ tel que $Eu_t^{4+\eta_3} < +\infty$.

B7 La matrice C_{α_0,β_0} est inversible.

L'hypothèse **B5** est une hypothèse technique nécessaire pour obtenir une convergence de $E\frac{1}{h_t}$ suffisament rapide vers 0. Avec ces hypothèses, on obtient le résultat suivant.

Théorème 1.6.4. Sous les hypothèses **B0-B4**, on obtient

$$\sqrt{n}\begin{pmatrix} \alpha_n - \alpha_0 \\ \beta_n - \beta_0 \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, C_{\alpha_0,\beta_0}^{-1} A_{\alpha_0,\beta_0} C_{\alpha_0,\beta_0}^{-1}\right),$$

où on aura définit pour $i, j \in \{2, 3\}$

$$D_t^{\theta_i,\theta_j} = \frac{\partial v_t}{\partial\theta_i\partial\theta_j}\left(1 - \frac{u_t^2}{v_t}\right) + \frac{\partial v_t}{\partial\theta_i}\frac{\partial v_t}{\partial\theta_j}\left(2\frac{u_t^2}{v_t} - 1\right)$$

$$C_{\alpha,\beta} = E\begin{pmatrix} D_1^{\alpha,\alpha} & D_1^{\alpha,\beta} \\ D_1^{\alpha,\beta} & D_1^{\beta,\beta} \end{pmatrix},$$

ainsi que $A_{\alpha,\beta}$ défini par

$$A_{\alpha,\beta} = E\left[\left(1 - \frac{u_1^2}{v_1(\alpha_0,\beta_0)}\right)^2 \frac{1}{v_1}\frac{\partial v_1}{\partial(\alpha,\beta)'}\frac{1}{v_1}\frac{\partial v_1}{\partial(\alpha,\beta)}\right].$$

Il est intéressant de remarquer le cas particulier de l'estimation d'un modèle ARCH(1) sur un MS-ARCH(1), dans ce cas on n'a pas besoin de supposer qu'il existe une unique pseudo vraie valeur du paramètre. On peut prouver que la composante α_n de l'estimateur de quasi maximum de vraisemblance du modèle mal spécifié converge vers $\alpha_0 = E\left[\alpha(\Delta_t)\right]$. C'est le seul cas où on obtient une valeur explicite pour la pseudo vraie valeur vers laquelle converge l'estimateur du QML.

1.6.4 Expériences numériques

Cette partie consacrée aux simulations est particulièrement importante pour ce chapitre. En effet, dans le cas où l'on estime un modèle GARCH$(1,1)$ quand le DGP est un processus MS-GARCH stationnaire ou explosif, on sait que les estimateurs des paramètres convergent vers une pseudo vraie valeur, mais cette valeur reste à être déterminée. Dans le cas où le DGP est un modèle ARCH explosif, on connait la pseudo vraie valeur α_0. On commence donc par vérifier numériquement la convergence de l'estimateur α_n vers la valeur théorique α_0. On montre ensuite, que dans le cas ARCH, la pseudo vraie valeur (inconnue dans le cas stationnaire) ne dépend pas des probabilités de transition de la chaîne de Markov mais seulement de sa distribution marginale, c'est-à-dire des probabilités stationnaires de cette chaîne.

Nous étudions ensuite par simulation l'estimation d'un modèle GARCH$(1,1)$ mal spécifié lorsque le DGP est un modèle MS-GARCH$(1,1)$ où la chaîne de Markov $(\Delta_t)_t$ peut prendre ses valeurs dans deux états. Nous analysons la capacité de prédiction du modèle mal spécifié. Si le modèle GARCH$(1,1)$ ne correspond pas au modèle du DGP, est-il capable d'en expliquer une partie de la dynamique. Nous comparons ainsi les erreurs quadratiques moyennes de la prévision issue du modèle mal spécifié et de la prévision théorique obtenue en supposant que l'on observe la chaîne de Markov $(\Delta_t)_t$ ainsi que la séquence des $(h_t)_t$. Nous trouvons que dans de nombreux cas, ces rapports sont de l'ordre de 0.9, indiquant que le modèle mal spécifié conserve une forte capacité de prédiction. Lorsque les deux états sont extrêmement différents, la capacité de prédiction du modèle mal spécifié devient plus mauvaise. On étudie également ce rapport dans le cas où le DGP est un modèle MS-GARCH non stationnaire et explosif et nous obtenons des résultats similaires.

Bibliographie

E. Andreou and E. Ghysels. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics*, 17(5) :579–600, 2002.

B. Andrews, M. Calder, and R.A. Davis. Maximum likelihood estimation for α-stable autoregressive processes. *The Annals of Statistics*, 37(4) :1946–1982, 2009.

A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B) :4046–4087, 2009.

L. Bachelier. Théorie de la spéculation. In *Annales Scientifiques de l'École Normale Superieure*, pages 21–86, 1900.

B. Basrak, R.A. Davis, and T. Mikosch. Regular variation of GARCH processes. *Stochastic Processes and their Applications*, 99(1) :95–115, 2002.

S.K. Basu and M. Maejima. A local limit theorem for attractions under a stable law. 87(01) : 179–187, 1980.

L. Bauwens, S. Laurent, and J.V.K. Rombouts. Multivariate GARCH models : a survey. *Journal of applied econometrics*, 21(1) :79–109, 2006.

I. Berkes and L. Horváth. The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics*, 32(2) :633–655, 2004.

I. Berkes, L. Horvath, and P. Kokoszka. GARCH processes : structure and estimation. *Bernoulli*, pages 201–227, 2003.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3) :307–327, 1986.

T. Bollerslev. Glossary to ARCH (GARCH). *CREATES Research Paper*, 49, 2008.

P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52(1-2) :115–127, 1992. ISSN 0304-4076.

F. Boussama. Ergodicité, mélange et estimation dans les modèles GARCH. Doctoral thesis, Université Paris 7. 1998.

F. Comte and O. Lieberman. Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, 84(1) :61–84, 2003.

R. Dahlhaus and W. Wefelmeyer. Asymptotically optimal estimation in misspecified time series models. *The Annals of Statistics*, 24(3) :952–974, 1996.

R.A. Davis and T. Mikosch. Extreme value theory for GARCH processes. *Handbook of Financial Time Series*, pages 187–200, 2009a.

R.A. Davis and T. Mikosch. The extremogram : A correlogram for extreme events. *Bernoulli*, 15 (4) :977–1009, 2009b.

I. Domowitz and H. White. Misspecified models with dependent observations. *Journal of Econometrics*, 20(1) :35–58, 1982.

R. Douc and E. Moulines. Asymptotic properties of the maximum likelihood estimation in misspecified Hidden Markov models. *Arxiv preprint arXiv :1110.0356*, 2011.

R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, pages 987–1007, 1982.

E.F. Fama. The behavior of stock-market prices. *The journal of Business*, 38(1) :34–105, 1965.

C. Francq and J.M. Zakoïan. The L2-structures of standard and switching-regime GARCH models. *Stochastic processes and their applications*, 115(9) :1557–1582, 2005.

C. Francq and J. Zakoian. Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theory*, 22(5) :815, 2006.

C. Francq and J.M. Zakoïan. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4) :605–638, 2004.

C. Francq and J.M. Zakoian. Optimal predictions of powers of conditionally heteroskedastic processes. MPRA Paper 22155, University Library of Munich, Germany, April 2010. URL http://ideas.repec.org/p/pra/mprapa/22155.html.

P. Fryzlewicz and S. Subba Rao. Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli*, 17(1) :320–346, 2011.

E. Ghysels, A.C. Harvey, and E. Renault. Stochastic volatility. *Handbook of Statistics*, 14 : 119–191, 1996.

B.V. Gnedenko, A.N. Kolmogorov, K.L. Chung, and J.L. Doob. *Limit distributions for sums of independent random variables.* Addison-Wesley Reading, MA, 1968.

C. Gourieroux, A. Monfort, and A. Trognon. Pseudo Maximum Likelihood Methods : Theory. *Econometrica*, 52(3) :681–700, 1984.

P. Hall and Q. Yao. Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, pages 285–317, 2003.

A.G. Halunga and C.D. Orme. First-Order Asymptotic Theory For Parametric Misspecification Tests Of GARCH Models. *Econometric Theory*, 25(2) :364, 2010.

E. Hillebrand. Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129 (1) :121–138, 2005.

S. Hörmann. Augmented GARCH sequences : Dependence structure and asymptotics. *Bernoulli*, 14(2) :543–561, 2008.

L. Hu and Y. Shin. Optimal test for Markov switching GARCH models. *Studies in Nonlinear Dynamics & Econometrics*, 12(3) :1528–1528, 2008.

A.T. Jensen and T. Lange. On convergence of the QMLE for misspecified GARCH models. *Journal of Time Series Econometrics*, 2(1) :1–29, 2010.

S.W. Lee and B.E. Hansen. Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. *Econometric theory*, 10 :29–29, 1994.

E. Liebscher. Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis*, 26(5) :669–689, 2005.

J.C. Liu. Stationarity of a Markov-switching GARCH model. *Journal of Financial Econometrics*, 4(4) :573–593, 2006.

S.M. Liu and B.W. Brorsen. Maximum likelihood estimation of a GARCH-stable model. *Journal of Applied Econometrics*, 10(3) :273–285, 1995.

A.W. Lo and A. Craig MacKinlay. An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1-2) :181–211, 1990.

R.L. Lumsdaine. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1, 1) and covariance stationary GARCH (1, 1) models. *Econometrica*, 64(3) : 575–596, 1996.

S. Lundbergh and T. Teräsvirta. Evaluating GARCH models. *Journal of Econometrics*, 110(2) : 417–435, 2002.

B. Mandelbrot. The variation of certain speculative prices. *The journal of business*, 36(4) : 394–419, 1963. ISSN 0021-9398.

M. Meitz and P. Saikkonen. Ergodicityn mixing and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory*, 24(05) :1291–1320, 2008.

T. Mikosch, G. Samorodnitsky, and L. Tafakori. Fractional moments of solutions to stochastic recurrence equations. 2012. Working paper.

S. Mittnik, M.S. Paolella, and S.T. Rachev. Stationarity of stable power-GARCH processes. *Journal of Econometrics*, 106(1) :97–107, 2002.

D.B. Nelson. Stationarity and Persistence in the GARCH (1, 1) Model. *Econometric Theory*, 6 (03) :318–334, 1990.

D.B. Nelson. Conditional heteroskedasticity in asset returns : a new approach. *Econometrica : Journal of the Econometric Society*, pages 347–370, 1991.

D.B. Nelson. Filtering and forecasting with misspecified ARCH models I : : Getting the right variance with the wrong model. *Journal of Econometrics*, 52(1-2) :61–90, 1992.

D.B. Nelson and D.P. Foster. Filtering and forecasting with misspecified ARCH models II : : Making the right forecast with the wrong model. *Journal of Econometrics*, 67(2) :303–335, 1995.

L. Qi, D. Xiu, and J. Fan. Non-Gaussian Quasi Maximum Likelihood Estimation of GARCH Models. *Arxiv preprint arXiv :1001.3895*, 2010.

D. Straumann. *Estimation in conditionally heteroscedastic time series models*, volume 181. Springer Verlag, 2005.

SJ Taylor. Asset price dynamics, volatility and prediction. 2005.

A.A. Weiss. Asymptotic Theory for ARCH Models : Estimation and Testing. *Econometric Theory*, 2(01) :107–131, 1986.

H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, pages 1–25, 1982.

H. White. Maximum likelihood estimation of misspecified dynamic models. *Misspecification Analysis, New York : Springer*, pages 1–19, 1984.

J.M. Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18 (5) :931–955, 1994.

K. Zhu and S. Ling. Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA–GARCH/IGARCH models. *The Annals of Statistics*, 39(4) :2131–2163, 2011.

V.M. Zolotarev. *One-dimensional stable distributions.* American Mathematical Society, 1986. ISBN 0821845195.

Chapter 2

Two-stage non Gaussian QML estimation of GARCH Models and testing the efficiency of the Gaussian QMLE

Cet article, coécrit avec Christian Francq et Jean-Michel Zakoïan a été publié dans la revue *Journal of Econometrics*, volume 165, page 246–257 en 2011.

2.1 Introduction

Since the introduction of ARCH models by Engle (1982) and their generalization by Bollerslev (1986), numerous papers studied the asymptotic distribution of the least-squares (LS) and the *Gaussian* quasi-maximum likelihood estimator (QMLE) of the GARCH parameters. Articles of the first generation focused on deriving the asymptotic distributions of these estimators. References along these lines are, among others, Weiss (1986), Lee and Hansen (1994), Lumsdaine (1996). A second generation of papers were concerned about reducing the assumptions required for the validity of the asymptotic properties of such estimators. Berkes et al. (2003) and Francq and Zakoïan (2004) obtained minimal assumptions ensuring the consistency and asymptotic normality (CAN) of the QMLE when the innovations have a finite fourth-order moment and under the strict stationarity condition; Hall and Yao (2003) derived the asymptotic distribution of the QMLE when the innovations have an infinite variance, under a restriction of the parameter space; Ling (2007) proposed a self-weighted QMLE and showed that it is CAN under only a fractional moment condition on the innovation process; Bardet and Wintenberger (2009) proved the asymptotic properties of the Gaussian QMLE for a general class of multidimensional causal processes.

An important issue is the possible efficiency loss of the QMLE, resulting from the use of an inappropriate Gaussian error distribution. In practice, the true error distribution is of course unknown and the MLE cannot be computed. Berkes and Horváth (2004) considered a non Gaussian QML framework. While the consistency of the Gaussian QMLE requires that the errors have a unit variance, Berkes and Horváth showed that non Gaussian QMLE can be CAN under alternative identifiability assumptions.

The main goal of the present paper is to study whether efficiency gains can be reached from using generalized QML estimators. To this aim, we introduce a class of densities (h_r) indexed by a positive parameter r, including the Gaussian density for $r = 2$, and we define QMLE's based on such densities. Such estimators being consistent under the assumption that the absolute errors to the power r have unit expectation, a reparameterization of the model is required. More precisely, under standard notation, we exploit the fact that the standard GARCH model

$$\epsilon_t = \sigma_t(\theta_0)\eta_t, \qquad E\eta_t^2 = 1$$

can be equivalently rewritten as

$$\epsilon_t = \sigma_t(\theta_0^{(r)})\eta_t^{(r)}, \qquad E|\eta_t^{(r)}|^r = 1,$$

where $\eta_t^{(r)} = \eta_t/\{E|\eta_t|^r\}^{1/r}$, provided $E|\eta_t|^r < \infty$. The initial parameter θ_0 is related to $\theta_0^{(r)}$ by

$$\theta_0 = B^{(r)}\theta_0^{(r)}, \tag{2.1.1}$$

where $B^{(r)}$ is a matrix which only depends on $E|\eta_t|^r$. Under appropriate conditions, a non Gaussian QMLE based on h_r allows to consistently estimate $\theta_0^{(r)}$. An estimator of $B^{(r)}$ can then be constructed using the standardized residuals $\widehat{\eta}_t^{(r)}$. From this we deduce a two-step estimator of θ_0 based on (2.1.1), which we call 2QMLE. Similarly to two-stage LSE, or Quasi-Generalized LSE, this estimator uses residuals from a first-step estimation to improve the asymptotic efficiency.

Indeed, it turns out that, when $r \leq 2$, the new estimator (i) is CAN under the same assumptions as the Gaussian QMLE, (ii) can be asymptotically more efficient than the QMLE.

The latter property depends on the distribution of η_t, not on θ_0. From this condition, for any error distribution it is possible to define an "optimal" power r_0, as the minimizer of the asymptotic variance of the QMLE. The second aim of this paper is to make an inference about this optimal power. We propose a consistent estimator of r_0 and derive its asymptotic distribution. An important consequence will be the possibility to test for the optimality of any value of r, in particular $r = 2$. In other words, we suggest a method for deciding whether the Gaussian QMLE can be considered as the most efficient QMLE for a given series. The whole procedure is simple to implement.

The present paper is related to Berkes and Horváth (2004), who studied the asymptotic behavior of non Gaussian QMLE for estimating reparameterizations of the GARCH model, and to Francq and Zakoian (2010), who used non Gaussian QMLE for predicting powers of GARCH models. Mukherjee (2006) also uses non Gaussian QMLE for estimating ARCH models identified by imposing an intercept equal to one. By contrast with these references, the focus of the present paper is on estimating the initial (standard) parameterization, and to develop a test for determining the most adequate QML.

The estimation method of this paper has common objectives with adaptive GARCH estimation methods. Both procedures rely on the standard QMLE in a first step, and use the standardized residuals to estimate characteristics of the noise distribution, in order to get efficiency gains in a second step. A difference, however, is that adaptive methods aim at reaching the asymptotic efficiency of the MLE, while our method aims at improving the asymptotic accuracy of the Gaussian QMLE. As shown in Drost and Klaassen (1997) (see also Ling (2003), and Ling and McAleer (2003) it is not possible to adaptively estimate all the GARCH parameters. After a suitable reparametrization, it is however possible to adaptively estimate some of them. Semiparametric two-step estimators, also considered by Engle and Gonzalez-Rivera (1991) or more recently by Di and Gangopadhyay (2011), can also be viewed as a sieve MLE. See Gallant and Nychka (1987) for semi-non parametric MLE in econometric models, and Chen (2007) for a review on sieve estimation in semi-non parametric models.

The plan of the rest of the paper is as follows. In Section 2.2 we define a generalized QMLE for a reparameterization of the GARCH model and we deduce the 2QMLE. We derive the asymptotic properties of the 2QMLE and compare its asymptotic efficiency with that of the Gaussian QMLE. Section 2.3 is devoted to the estimation of the optimal power r_0, and to the test of the hypothesis $r_0 = 2$. Simulation experiments are presented in Section 2.4. In particular, our approach will be compared with adaptive estimation methods. An empirical application based on major stock indices is presented in Section 2.5. Section 2.6 proposes some concluding remarks. Proofs are deferred to Section 2.7.

2.2 Two-stage QML estimation

The model we consider in this paper is the standard GARCH(p, q) defined by

$$\begin{cases} \epsilon_t = \sigma_{0t}\eta_t \\ \sigma_{0t}^2 = \omega_0 + \sum_{i=1}^{q} \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j}\sigma_{0t-j}^2 \end{cases} \tag{2.2.1}$$

where (η_t) is a sequence of independent and identically distributed random variables, with $E\eta_t = 0$ and

$$E\eta_t^2 = 1,$$

η_t being independent of $\{\epsilon_u, u < t\}$. It is assumed that

$$\theta_0 = (\omega_0, \alpha_{01}, \dots, \alpha_{0q}, \beta_{01}, \dots, \beta_{0p})' \in (0, +\infty) \times [0, \infty)^{p+q}$$

Let the polynomials $\mathcal{A}_\theta(z) = \sum_{i=1}^q \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^p \beta_j z^j$ where

$$\theta = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)'.$$

For θ such that $\sum_{j=1}^p \beta_j < 1$ and $\beta_j \geq 0$, for $j = 1, \dots, p$ define the function $\sigma_t^2(\theta) = \frac{\omega_0}{\mathcal{B}_\theta(1)} + \mathcal{B}_\theta^{-1}(L)\mathcal{A}_\theta(L)\epsilon_t^2$, where L denotes the lag operator. We have $\sigma_{0t}^2 = \sigma_t^2(\theta_0)$.

Next, we consider the estimation of a reparameterization of the standard GARCH Model (2.2.1) under an alternative identifiability condition.

2.2.1 Generalized QMLE of a reparameterized model

For any $r > 0$ such that $E|\eta_t|^r < \infty$, the GARCH(p, q) model can be equivalently rewritten as

$$\begin{cases} \epsilon_t = \sigma_{0t}^{(r)} \eta_t^{(r)} \\ \sigma_{0t}^{(r)2} = \omega_0^{(r)} + \sum_{i=1}^q \alpha_{0i}^{(r)} \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j}^{(r)} \sigma_{0t-j}^{(r)2} \end{cases}$$

with

$$E\left|\eta_t^{(r)}\right|^r = 1, \tag{2.2.2}$$

where $\theta_0^{(r)} = (\omega_0^{(r)}, \alpha_{01}^{(r)}, \dots, \alpha_{0q}^{(r)}, \beta_{01}^{(r)}, \dots, \beta_{0p}^{(r)})'$ is the true parameter value in this reparameterization. Let $\mu_s = E|\eta_t|^s$ and $\mu_s^{(r)} = E|\eta_t^{(r)}|^s$ for any $s > 0$, provided these moments exist. We have $\sigma_{0t}^{(r)2} = \sigma_t^2(\theta_0^{(r)}) = \mu_r^{2/r}\sigma_{0t}^2$, $\eta_t^{(r)} = \mu_r^{-1/r}\eta_t$ and (2.1.1) holds with

$$B^{(r)} = \begin{pmatrix} \mu_r^{-2/r} I_{q+1} & 0 \\ 0 & I_p \end{pmatrix} = \begin{pmatrix} \mu_2^{(r)} I_{q+1} & 0 \\ 0 & I_p \end{pmatrix}.$$

Notice that the last GARCH coefficients are unchanged in this reparameterization, namely $\beta_{0j}^{(r)} = \beta_{0j}$ for all j.

For some parameter space $\Theta^{(r)}$, for $\theta = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p) \in \Theta^{(r)}$, and, given initial values $\epsilon_0, \cdots, \epsilon_{1-q}, \tilde{\sigma}_0^2(\theta), \cdots, \tilde{\sigma}_{1-p}^2(\theta)$, we define positive variables $\tilde{\sigma}_t(\theta)$ via the recursion

$$\tilde{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \tilde{\sigma}_{t-j}^2(\theta), \qquad t \geq 1.$$

For example, we can take $\epsilon_0 = \cdots = \epsilon_{1-q} = \tilde{\sigma}_0^2(\theta) = \cdots = \tilde{\sigma}_{1-p}^2(\theta) = \omega$. For any density f, a generalized QMLE is defined as any minimizer over $\Theta^{(r)}$ of

$$\frac{-1}{n} \sum_{t=1}^n \log \left\{ \frac{1}{\tilde{\sigma}_t(\theta)} f\left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)} \right) \right\}.$$

Imposing an identifiability condition of the form (2.2.2) reduces the choice of the possible densities f. Francq and Zakoian (2010) showed that to obtain a consistent estimator of $\theta_0^{(r)}$, it is necessary to choose $f = h_r$ where

$$h_r(x) = \frac{r^{1-1/r}}{2\Gamma(1/r)} \exp\left(-\frac{1}{r}|x|^r\right).$$

This density is called a generalized Gaussian density (also known as the Generalized Error Distribution (GED(r)) or the power Gamma distribution). Note that h_2 is the standardized Gaussian density, which is used to estimate $\theta_0^{(2)} = \theta_0$. We consider the generalized QMLE of $\theta_0^{(r)}$,

$$\widehat{\theta}_n^{(r)} = \underset{\theta \in \Theta^{(r)}}{\operatorname{argmin}}\ \tilde{\mathbf{I}}_n(\theta),$$

where for $\theta \in \Theta^{(r)}$, and for some constant K,

$$\tilde{\mathbf{I}}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta)$$

with

$$\tilde{l}_t(\theta) = \log \tilde{\sigma}_t^2(\theta) + \frac{2}{r}\frac{|\epsilon_t|^r}{\tilde{\sigma}_t^r(\theta)} = -2\log\left(\frac{1}{\tilde{\sigma}_t(\theta)} h_r\left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}\right)\right) + K.$$

2.2.2 Asymptotic properties of the two-stage estimator of θ_0

We now turn to the estimation of the parameter of interest, θ_0. We rely on (2.1.1), using the generalized QMLE $\widehat{\theta}_n^{(r)}$ of $\theta_0^{(r)}$ and the empirical counterpart of the second-order moment of $\eta_t^{(r)}$. To estimate the matrix $B^{(r)}$, we define the standardized residuals

$$\widehat{\eta}_t^{(r)} = \frac{\epsilon_t}{\tilde{\sigma}_t(\widehat{\theta}_n^{(r)})}, \qquad t = 1, \dots, n.$$

Let for any $s > 0$,

$$\widehat{\mu}_{s,n}^{(r)} = \frac{1}{n}\sum_{t=1}^{n} \left|\widehat{\eta}_t^{(r)}\right|^s,$$

and let

$$\widehat{B}_n^{(r)} = \begin{pmatrix} \widehat{\mu}_{2,n}^{(r)} I_{q+1} & 0 \\ 0 & I_p \end{pmatrix}.$$

Let $\widehat{\theta}_{n,r}$ be the *two-stage* QMLE (2QMLE) of θ_0 defined as

$$\widehat{\theta}_{n,r} = \widehat{B}_n^{(r)} \widehat{\theta}_n^{(r)}. \tag{2.2.3}$$

The next result establishes the asymptotic properties of this estimator. Let $\Theta = \{B^{(r)}\theta, \quad \theta \in \Theta^{(r)}\}$. Denote by $\gamma(A_0)$ the top-Lyapounov exponent associated to Model (2.2.1) (see for instance Berkes and Horváth (2004). The following assumptions will be required for the CAN of $\widehat{\theta}_{n,r}$.

A1 $\theta_0 \in \Theta$ and Θ is a compact.

A2 $\gamma(A_0) < 0$ and $\forall \theta \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$.

A3 $\mu_2 = 1$, $\mu_r < \infty$ and $|\eta_t|$ takes at least 5 different values.

A4 For $p > 0$, the polynomials $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ have no common roots, $\mathcal{A}_{\theta_0}(1) \neq 0$ and $\alpha_{0q} + \beta_{0p} \neq 0$.

A5 $\mu_4 < \infty$ and $\mu_{2r} < \infty$.

A6 θ_0 belongs to the interior of Θ.

Note that the assumptions made to ensure the CAN of the 2QMLE are almost the same as those made by Francq and Zakoïan (2004) for the Gaussian QMLE. When $r \leq 2$, only the identifiability assumption A3 is slightly stronger.

Theorem 2.2.1. *Let $r > 0$. Under Assumptions* **A1-A4**, *the 2QMLE of θ_0 satisfies*

$$\widehat{\theta}_{n,r} \underset{n\to+\infty}{\longrightarrow} \theta_0, \quad a.s.$$

If, in addition, **A5-A6** *hold,*

$$\sqrt{n}\left(\widehat{\theta}_{n,r} - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_r) \tag{2.2.4}$$

with

$$\Sigma_r = g(r)J^{-1} + \{\mu_4 - 1 - g(r)\}\, \overline{\theta}_0\overline{\theta}_0', \qquad g(r) = \left(\frac{2}{r}\right)^2 \left(\frac{\mu_{2r}}{\mu_r^2} - 1\right),$$

and

$$\overline{\theta}_0 = (\omega_0, \alpha_{01}, \dots, \alpha_{0q}, 0, \dots, 0)', \quad J = E\left(\phi_t\phi_t'\right),$$
$$\phi_t = \phi_t(\theta_0), \phi_t(\theta) = \frac{1}{\sigma_t^2(\theta)} \frac{\partial \sigma_t^2}{\partial \theta}(\theta).$$

Remark 2.2.1. It can be shown that, if the initial values are chosen equal to zero and if the ϵ_t's are not all equal to zero, we have $\hat{\mu}_{2,n}^{(2)} = 1$. In this case, the estimator $\widehat{\theta}_{n,2}$ coincides with the Gaussian QMLE. If other initial values are chosen, the estimators may not coincide exactly but their asymptotic distributions are the same.

The next result shows that the comparison of the asymptotic efficiencies of the standard QMLE and the 2QMLE only depends on a scalar function of r, and does not depend on θ_0.

Corollary 2.2.1. *The asymptotic efficiency of $\widehat{\theta}_{n,r}$ decreases when $g(r)$ increases, and the estimator $\widehat{\theta}_{n,r}$ is asymptotically more efficient than $\widehat{\theta}_{n,2}$ iff*

$$g(r) < \mu_4 - 1. \tag{2.2.5}$$

Moreover

$$\Sigma_r - \Sigma_2 = \{g(r) - \mu_4 + 1\}(J^{-1} - \overline{\theta}_0\overline{\theta}_0').$$

Remark 2.2.2. The quantities $g(r)$ and $\mu_4 - 1$ are unknown, but they can be easily estimated by empirical moments of the standardized residuals of the Gaussian QMLE. As will be seen in Section 2.3, this can be used to test whether the 2QMLE is more efficient than the Gaussian QMLE or not.

2.2.3 Illustrations

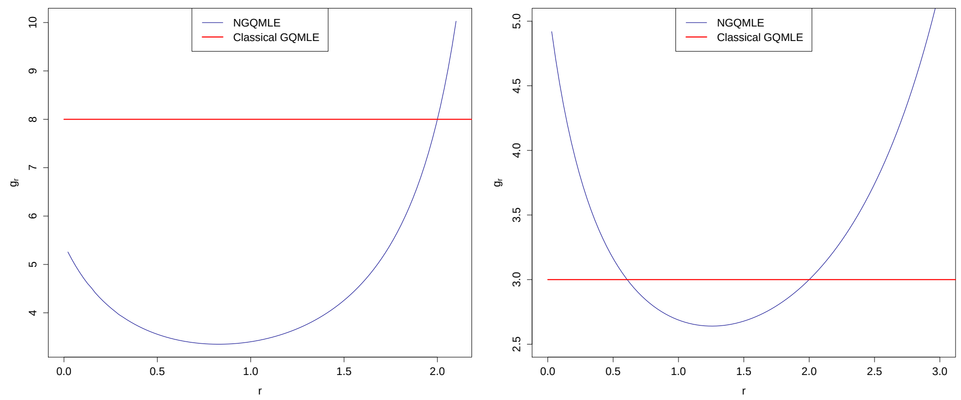
On different examples of classical distributions for η_t, we determine the range of powers r for which (2.2.5) is satisfied.

Example 2.2.1. (Standard Gaussian distribution) Since the Gaussian QML method is equivalent to the exact ML method, there is no value of r that verifies $g(r) < \mu_4 - 1$ when $f \sim \mathcal{N}(0,1)$. For $r = 2$ we find $g(2) = \mu_4 - 1$. Figure 2.1, displaying the graph of g as a function of r, shows that the efficiency loss of the 2QMLE with respect to the QMLE increases when r goes to zero or to infinity.

Figure 2.1: Function g (in blue) and $\mu_4 - 1$ (in red) for the standard Gaussian Distribution

Example 2.2.2. (Student distributions) When the distribution of (η_t) is a Student with ν degrees of freedom (standardized so that $E\eta_t^2 = 1$), $g(r)$ is only defined for $r \in \left(0, \frac{\nu}{2}\right)$ and is given by

$$g(r) = \frac{4}{r^2}\left(\frac{\sqrt{\pi}\Gamma\left(\frac{1}{2}+r\right)\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}(-2r+\nu)\right)}{\Gamma\left(\frac{1+r}{2}\right)^2\Gamma\left(\frac{\nu-r}{2}\right)^2} - 1\right), \quad r \in \left(0, \frac{\nu}{2}\right).$$



(a) $\nu = 5$ (b) $\nu = 10$

Figure 2.2: Function g (in blue) and $\mu_4 - 1$ (in red) for Student Distributions

In Figure 2.2, it is seen that the standard Gaussian QMLE outperforms the 2QMLE for large values of r, and also for small values of r when ν is large. However, it is interesting to see that the Gaussian QMLE can be asymptotically less efficient than the 2QMLE estimators for $r < 2$. Note also the existence of an "optimal" value r_0 of r, which minimizes g over the real line. Figure 2.3 shows r_0 (obtained numerically) as a function of ν. When ν tends to infinity, the Student distribution converges to the Gaussian distribution, and r_0 tends to 2.

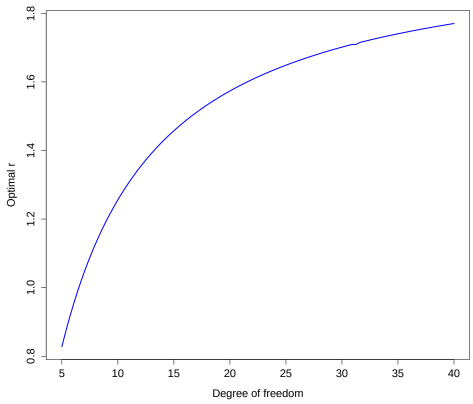


Figure 2.3: Optimal power r_0 as a function of ν for Student distributions

Example 2.2.3. (Uniform distribution) When (η_t) follows the uniform distribution on $\left[-\sqrt{3}, \sqrt{3}\right]$, we have

$$g(r) = \frac{4}{r^2}\left(\frac{(1+r)^2}{1+2r} - 1\right).$$

Contrary to the previous examples, Figure 2.4 shows that the standard QML is outperformed for large values of r. Notice that no optimal r_0 exists in this case.

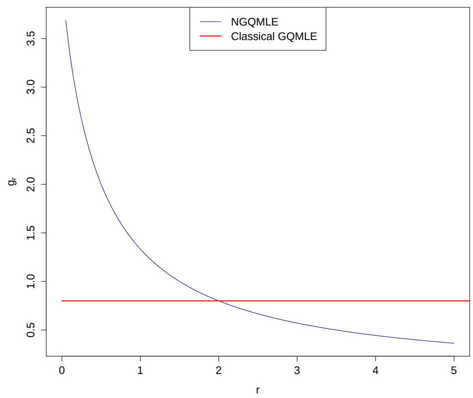


Figure 2.4: Function g (in blue) and $\mu_4 - 1$ (in red) for the uniform distribution

Generally for all bounded distribution, $g(r)$ will tend to 0 when $r \to +\infty$.

2.2.4 Optimal 2QMLE

As illustrated by the previous examples, for most standard distributions, a positive value of r minimizing the asymptotic variance of the 2QMLE, that is minimizing g, exists. We therefore introduce the following assumption. Let $R = [\underline{r}, \overline{r}] \subset (0, r_{\max})$ where $r_{\max} = \sup\{r \in \mathbb{R}\ ;\ \mu_{2r} < \infty\}$.

A7 There exists a unique $r_0 > 0$ such that $r_0 = \underset{r \in R}{\operatorname{argmin}}\ g(r)$.

We call *optimal* 2QMLE the "estimator" $\widehat{\theta}_{n,r_0}$. Note that, the distribution of η_t being generally unknown, this estimator cannot be used in practical situations. For practical purposes, we will define a *feasible* optimal 2QMLE, based on a consistent estimator $\widehat{r}_n$ of r_0.

2.3 Estimating and testing for r_0

In this section, we consider estimating and testing r_0. Because $g(r)$ only depends on the distribution of η_t it can be estimated, for any value of r, using standardized residuals $\widehat{\eta}_t$ obtained from a Gaussian QML estimation. An estimator of r_0 can then be obtained by minimizing the estimator of g.

For u such as $\mu_u < \infty$, let

$$\widehat{\mu}_{n,u} = \frac{1}{n}\sum_{t=1}^{n} |\widehat{\eta}_t|^u \quad \text{with} \quad \widehat{\eta}_t = \frac{\epsilon_t}{\tilde{\sigma}_t\left(\widehat{\theta}_{n,2}\right)},$$

let the estimator of $g(r)$

$$\widehat{g}_n(r) = \left(\frac{2}{r}\right)^2 \left(\frac{\widehat{\mu}_{n,2r}}{\widehat{\mu}^2_{n,r}} - 1\right),$$

and let

$$\widehat{r}_n = \underset{r\in R}{\text{argmin}}\ \widehat{g}_n\left(r\right). \tag{2.3.1}$$

The next result gives the asymptotic distribution of $\widehat{r}_n$.

Theorem 2.3.1. *Under Assumptions* **A1-A4** *and* **A7***, and if* $E\left\{|\eta_t|^{4r_0}\left(\log|\eta_t|\right)^2\right\} < \infty$*, we have* $\widehat{r}_n \to r_0$ *a.s. and if* r_0 *belongs to the interior of* R*,*

$$\sqrt{n}\left(\widehat{r}_n - r_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \tau_{r_0}\right) \tag{2.3.2}$$

where τ_{r_0}*, whose expression is given in (2.7.21) below, only depends on the distribution of* η_t*.*

Remark 2.3.1. It is interesting to note that the estimation of θ_0 has no effect on the asymptotic distribution of $\widehat{r}_n$. In other words, the asymptotic distribution of $\widetilde{r}_n = \underset{r\in R}{\text{argmin}}\ \widetilde{g}_n\left(r\right)$, where $\widetilde{g}_n$ is defined as $\widehat{g}_n$ but with the $\widehat{\eta}_t$ replaced by the η_t, is also given by (2.3.2). The reason for this surprising property is the following: noting that $\eta_t = \epsilon_t/\sigma_t(\theta_0)$, the function $g(r)$ can be written as $g(r,\theta_0)$; in the proof of Lemma 2.7.10 below we will show that $\frac{\partial^2 g}{\partial r \partial \theta}(r_0,\theta_0) = 0$.

Remark 2.3.2. By continuity of $g(\cdot)$ and compactness of R, r_0 exists but is not necessarily unique. When the minimizer r_0 is not unique (that is when Assumption **A7** is not satisfied), a careful examination of the proof of Theorem 2.3.1 shows that r_n converges to the set $\{\underset{r\in R}{\text{Argmin}}\ g(r)\}$. For each element r of this set, Theorem 2.2.1 shows that the estimator $\widehat{\theta}_{n,r}$ has the same (optimal) asymptotic Gaussian distribution.

Remark 2.3.3. Let us discuss the existence of a minimizer of the function $g(\cdot)$ on $(0,\infty)$. When the distribution of η_1 does not admit moments of all orders, *i.e.* when $r_{\max} < \infty$ as in Example 2.2.2, then any minimizer of $g(\cdot)$ is less than $r_{\max}$. The Gaussian case, considered in Example

2.2.1, reveals that $g(\cdot)$ may also admit a minimizer in $(0, \infty)$ when η_t admits moments of all orders. Finally, $g(r) \to 0$ as $r \to \infty$ for any distribution with compact support (as in Example 2.2.3). In this case a minimizer over $(0, \infty)$ may not exist.

The following result shows that for the consistency of $\widehat{r}_n$ it is essential to perform the optimization (2.3.1) over a bounded set R.

Lemma 2.3.1. *If the ϵ_t, for $t = 1, \dots, n$, are not all equal to zero,*

$$\widehat{g}_n(r) \to 0, \quad a.s. \text{ as } r \to \infty.$$

The Gaussian QMLE being the most widely used estimator in GARCH models, it is of interest to test its optimality in the family of the estimators $\widehat{\theta}_{n,r}$. This leads to testing the hypothesis $r_0 = 2$. Note that the asymptotic variance τ_{r_0} of $\widehat{r}_n$ is a function of θ_0, r_0 and moments of η_t (see Equations (2.7.9), (2.7.17), (2.7.20) and (2.7.21) below). Let $\widehat{\tau}_2$ be the estimator obtained by replacing θ_0 by $\widehat{\theta}_{n,2}$, r_0 by 2 and all the theoretical moments of η_t by the corresponding empirical moments of the $\widehat{\eta}_t$'s in the expression of τ_{r_0}.

Corollary 2.3.1. *Under the assumptions of Theorem 2.3.1 and under $H_0 : r_0 = 2$,*

$$\widehat{\zeta}_n = \frac{n}{\widehat{\tau}_2}(\widehat{r}_n - 2)^2 \xrightarrow{\mathcal{L}} \chi^2(1).$$

To conclude this section, we propose the following scheme for implementing the 2QML method.[1]

- Step 1: compute the standard QMLE $\widehat{\theta}_{n,2}$ and $\widehat{r}_n$ given by (2.3.1);
- Step 2: compute the statistic ζ_n where $\widehat{\tau}_2$ is obtained by replacing theoretical moments by empirical ones in (2.7.21). For a given significance level $\underline{\alpha}$, if $\widehat{\zeta}_n$ is less than the $1-\underline{\alpha}$-quantile of the $\chi^2(1)$ distribution, the standard QMLE cannot be rejected. Otherwise, the standard QMLE is rejected and the feasible optimal 2QMLE based on $\widehat{\theta}_{n,r_n}$, using (2.2.3), should do a better job.

Our approach has apparent similarities with two-step statistical procedures in which model selection precedes estimation. Leeb and Pötscher (2006, 2009) show that such procedures may have poor asymptotic properties. Specifically, post-model-selection estimators cannot have satisfactory performance uniformly in the parameter space, even asymptotically. Our framework is different, in the sense that our first step does not involve model selection but rather, the choice of an appropriate estimation method. More precisely, for any given r (even far from r_0) the resulting estimator of θ_0 will remain CAN. However, the asymptotic distribution of the estimator of θ_0 is likely to be affected by a (consistent) data-chosen value of r. This issue is considered through simulations in the next section, theoretical investigations being deferred to future work.

2.4 Simulation experiments

In this section, we first study the finite sample performance of the test of optimality of the Gaussian QMLE. We then compare the feasible optimal 2QMLE to the standard Gaussian QMLE.

1. The R code used to implement this algorithm and all the numeric illustrations of the paper are available from the authors' web pages.

2.4.1 Finite sample performance of the test on r_0

Consider the test defined by Corollary 2.3.1, for testing $\mathbf{H_0} : r_0 = 2$, when $\eta_t \sim GED(r)$. For such a distribution, the minimizer of $g(r)$ is $r_0 = r$. We simulated $N = 1,000$ sample paths of the model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = 0.02 + 0.1\epsilon_{t-1}^2 + 0.8\sigma_{t-1}^2 \end{cases} \tag{2.4.1}$$

and estimated the parameters by Gaussian QML. For each sample path, the standardized residuals $\hat{\eta}_t$ are then used to compute the statistic $\widehat{\zeta}_n$. Table 2.1 reports the frequencies of rejection of $\mathbf{H_0} : r_0 = 2$. The test turns out to over-reject under the null, and to be slightly biased, even for large sample sizes.

$\alpha = 5\%$	$r = 1$	$r = 1.4$	$r = 1.7$	$r = 2$	$r = 2.3$	$r = 2.6$	$r = 3$
$n = 500$	0.90	0.68	0.52	0.46	0.43	0.48	0.58
$n = 1,000$	0.99	0.82	0.54	0.36	0.36	0.41	0.55
$n = 5,000$	1.00	0.99	0.78	0.18	0.14	0.43	0.70
$n = 10,000$	1.00	1.00	0.93	0.12	0.06	0.50	0.80

Table 2.1: Empirical size of the test of $\mathbf{H_0} : r_0 = 2$, when $\eta_t \sim GED(r)$.

In order to explain this finite sample bias, we replaced $\widehat{r}_n$ by $\widetilde{r}_n$ (*i.e.* $\hat{\eta}_t$ by η_t) in $\widehat{\zeta}_n$. In view of Remark 2.3.1, the asymptotic behavior of the test is not affected by this change. Table 2.2 shows that the finite sample behavior is however greatly improved. In Tables 2.1 and 2.2, one can notice an asymmetry around $r = 2$. Indeed, the test turns out to be more powerful for values of r smaller than 2 than for values of r greater than 2. This can be explained by the difficulty to estimate the asymptotic variance of $\widehat{r}_n$ for large values of r.

$\alpha = 5\%$	$r = 1$	$r = 1.4$	$r = 1.7$	$r = 2$	$r = 2.3$	$r = 2.6$	$r = 3$
$n = 500$	0.92	0.61	0.39	0.24	0.30	0.44	0.69
$n = 1,000$	0.98	0.78	0.41	0.17	0.27	0.51	0.79
$n = 5,000$	1.00	1.00	0.75	0.065	0.12	0.64	0.98
$n = 10,000$	1.00	1.00	0.92	**0.05**	0.06	0.74	1.00

Table 2.2: As in Table 2.1, but with $\widetilde{r}_n$ instead of the statistic $\widehat{r}_n$.

2.4.2 Finite-sample performance of the optimal 2QMLE's

We simulated $1,000$ samples of size $n \in \{200, 500, 1000, 10000\}$ of the GARCH$(1,1)$ model (2.4.1) for different distributions of η_t (Gaussian, GED(0.3), GED(1), $t_{2.5}$, t_{10}, a mixture of Gaussian and a mixture of t_5 with mean 2 and -2). All the distributions have been standardized in order to have zero mean and unit variance. For each distribution of η_t, we are able to compute (sometimes numerically) the theoretical value r_0 minimizing the asymptotic variance of the estimator $\widehat{\theta}_{n,r}$ defined by (2.2.3). For each sample, we computed three estimators: the Gaussian

QMLE $\widehat{\theta}_n^{(2)} \simeq \widehat{\theta}_{n,2}$, the unfeasible optimal 2QML estimator $\widehat{\theta}_{n,r_0}$, and the feasible optimal 2QML estimator $\widehat{\theta}_{n,\widehat{r}_n}$ with estimated r_0. If the null hypothesis $H_0: \; r_0 = 2$ cannot be rejected, then we do not compute $\widehat{\theta}_{n,\widehat{r}_n}$ but keep the value of $\widehat{\theta}_{n,2}$. Table 2.3 displays the relative efficiency (RE) of both the feasible and unfeasible 2QMLE with respect to the Gaussian QMLE, the RE being defined as the ratio of the MSE (mean squared errors) of the Gaussian QMLE with respect to the MSE of the alternative estimator. To obtain the non Gaussian QMLE, we need an estimator of the moment $\mu_2^{(r)}$. For large sample ($n = 10,000$), we used the empirical mean, but for smaller samples and especially in the case of fat tailed innovations ($t_{2.5}$ for example) we obtained better results with a trimmed estimator.

The first output of Table 2.3 is that $RE(\widehat{\theta}_{n,r_0})$ and $RE(\widehat{\theta}_{n,\widehat{r}_n})$ are generally close, which indicates that the feasible estimator $\widehat{\theta}_{n,\widehat{r}_n}$ enjoys similar properties to the unfeasible optimal estimator $\widehat{\theta}_{n,r_0}$. The second output is that $\widehat{\theta}_{n,\widehat{r}_n}$ and $\widehat{\theta}_{n,r_0}$ outperform the Gaussian QMLE in most cases (their RE are in general greater than one). Obviously, for the Gaussian distribution the optimal estimator is obtained for $r_0 = 2$, and thus $RE(\widehat{\theta}_{n,r_0})$ is very close to 1. Moreover, since the Gaussian QMLE is the exact maximum likelihood estimator, it cannot be asymptotically outperformed by another estimator. It is thus satisfactory to observe that, for $n \geq 1,000$, $\widehat{\theta}_{n,\widehat{r}_n}$ is in this case very close to the optimal MLE, the RE being only slightly less than 1. RE's greater than 1 can even be observed for small sample sizes. The most impressive improvements are obtained for the distributions with the largest tails, that is for the $t_{2.5}$ and the GED(0.3). For these distributions, the MSE of the 2QMLE can be over five times smaller than the MSE of the Gaussian QMLE. For some applications, it is relevant to consider the case where the error density is both fat tailed and multimodal (see *e.g.* Gallant, Rossi and Tauchen, 1992). Table 2.3 thus considers mixtures of distributions with two modes. As expected, the RE is still close to one when the tail is light (*i.e.* in the case of a balanced mixture of two Gaussian distributions), whereas the RE is in general greater than one when the tail is fatter (*i.e.* in the case of a balanced mixture of t_5).

		$RE(\widehat{\theta}_{n,\widehat{r}_n})$			$RE(\widehat{\theta}_{n,r_0})$		
	N	ω	α	β	ω	α	β
Gaussian	10000	0.89	0.77	0.98	1.00	0.99	1.00
	1000	1.06	1.01	0.98	1.05	1.03	1.00
	500	1.04	1.05	1.00	1.08	1.07	1.00
	200	0.94	1.02	0.81	1.17	1.13	1.00
GED(0.3)	10000	4.13	4.47	5.49	4.69	4.92	5.42
	1000	1.61	4.70	0.66	1.68	4.91	0.70
	500	2.04	3.23	0.43	2.18	3.42	0.45
	200	1.33	1.46	0.35	1.50	1.72	0.37
GED(1)	10000	1.12	0.99	1.46	1.34	1.47	1.47
	1000	1.41	0.81	1.24	1.39	0.81	1.18
	500	1.48	1.40	1.23	1.61	1.43	1.33
	200	0.96	1.35	0.72	1.02	1.43	0.70
Student $t_{2.5}$	10000	11.22	8.38	15.86	10.49	7.63	14.87
	1000	11.66	8.15	6.68	11.18	8.44	5.51
	500	5.46	4.57	1.43	4.58	5.70	1.15
	200	4.15	3.14	0.63	5.16	2.99	0.59
Student t_5	10000	1.60	1.46	1.70	1.74	1.67	1.71
	1000	1.73	0.91	1.73	1.73	0.91	1.78
	500	1.81	1.75	1.34	1.79	2.08	1.36
	200	1.47	1.78	1.04	1.42	1.95	0.99
Gaussian Mixture	10000	0.76	0.54	1.02	1.00	1.06	1.02
	1000	1.24	0.89	0.97	1.34	0.96	1.06
	500	1.10	1.02	1.07	1.16	1.10	1.10
	200	0.84	0.93	0.78	0.89	1.03	0.82
Student Mixture	10000	1.06	0.96	1.15	1.12	1.07	1.17
	1000	1.29	1.20	1.23	1.29	1.18	1.20
	500	1.19	1.09	1.07	1.19	1.10	1.07
	200	0.93	1.21	0.86	1.00	1.25	0.89

Table 2.3: Relative efficiency (RE) of the Generalized QMLEs with respect to the Gaussian QMLE.

Now we compare our approach with adaptive GARCH estimation. Following Engle and Gonzalez-Rivera (1991), in this approach the GARCH parameters are estimated by QMLE in a first step. An estimate $\hat{f}$ of the density f of η_t is obtained with a kernel estimator [2] based on $\widehat{\eta}_1, \dots, \widehat{\eta}_n$. In a second step, the semiparametric estimator is obtained by maximizing an approximation of the likelihood obtained by replacing the unknown density f by $\hat{f}$. Table 2.4 displays the relative efficiency of the feasible optimal 2QMLE with respect to that semiparametric estimator. The RE's are computed over 1000 samples of size 1000 of Model (2.4.1), for different distributions of η_t. Comparing Table 2.3 and Table 2.4, one can see that the semiparametric estimator generally improves upon the QMLE, particularly for heavy-tailed distributions, but

2. For the numerical results presented below, we used the Gaussian kernel, with a bandwidth equal to three times the default value of the R function `density()`.

that the improvement is in general less important than with the 2QMLE.

Distribution	ω	α	β
Gaussian	1.31	0.95	1.22
GED(0.3)	3.70	4.18	0.94
GED(1)	1.78	0.82	1.43
Student $t_{2.5}$	3.44	4.77	3.35
Student t_5	1.82	0.84	1.42
Gaussian Mixture	0.91	1.60	1.30
Student Mixture	0.81	1.19	0.99

Table 2.4: Relative efficiency of the 2QMLE with respect to the adaptive estimator.

2.5 Application

In this section, we consider daily returns of 10 indices, namely the CAC, DAX, DJA, DJI, DJT, DJU, FTSE, NIKKEI, SMI and SP500. A GARCH(1,1) model is estimated on each of these series. The samples extend from March 1, 1990 to March, 31, 2010. Table 2.5 displays the estimated parameters for the Gaussian QMLE and the feasible optimal 2QMLE. The parameters estimated by the two methods are close but the standard deviation estimated for the 2QML method are smaller for each indices. This is not surprising because, by construction, the power r_n minimizes the estimated variance of the 2QML.

Index	Estimator	ω	α	β
CAC	QMLE	0.031 (0.008)	0.087 (0.013)	0.897 (0.014)
	2QMLE	0.024 (0.003)	0.077 (0.007)	0.913 (0.009)
DAX	QMLE	0.049 (0.014)	0.099 (0.019)	0.875 (0.022)
	2QMLE	0.024 (0.004)	0.092 (0.009)	0.901 (0.010)
DJA	QMLE	0.015 (0.004)	0.079 (0.011)	0.908 (0.012)
	2QMLE	0.012 (0.002)	0.074 (0.006)	0.917 (0.009)
DJI	QMLE	0.012 (0.003)	0.075 (0.011)	0.916 (0.011)
	2QMLE	0.008 (0.001)	0.067 (0.006)	0.928 (0.007)
DJT	QMLE	0.030 (0.010)	0.077 (0.014)	0.911 (0.015)
	2QMLE	0.021 (0.003)	0.062 (0.006)	0.929 (0.008)
DJU	QMLE	0.021 (0.004)	0.111 (0.012)	0.871 (0.012)
	2QMLE	0.018 (0.002)	0.101 (0.007)	0.885 (0.010)
FTSE	QMLE	0.014 (0.003)	0.092 (0.010)	0.898 (0.010)
	2QMLE	0.013 (0.002)	0.085 (0.007)	0.905 (0.009)
NIKKEI	QMLE	0.058 (0.012)	0.105 (0.013)	0.874 (0.014)
	2QMLE	0.042 (0.006)	0.093 (0.007)	0.892 (0.011)
SMI	QMLE	0.054 (0.014)	0.129 (0.023)	0.829 (0.028)
	2QMLE	0.031 (0.004)	0.117 (0.011)	0.864 (0.013)
SP500	QMLE	0.009 (0.003)	0.072 (0.010)	0.921 (0.010)
	2QMLE	0.006 (0.001)	0.066 (0.006)	0.933 (0.007)

Table 2.5: QMLE and feasible optimal 2QMLE of GARCH(1,1) models for 10 daily stock market returns, from March 1, 1990 to March 31, 2010. The standard deviations are displayed in parentheses.

We now investigate whether the standard QMLE, corresponding to the hypothesis $\mathbf{H_0}: r_0 = 2$, is optimal for estimating these series. Table 2.6 displays the estimated variance minimizer r_n for the ten indices and the p-value of the test of Section 2.3 for the whole period, and for subperiods of two years. From Lemma 2.3.1 we know that the minimizer r_n systematically tends to infinity as the set R tends to $(0, \infty)$. For this reason we decided to declare the minimizer as not available (NA) when it reaches the upper boundary of the fixed compact R used in our algorithm.

The most striking feature is that the estimators of the optimal power r_0 are, in most cases, much smaller than 2. The assumption $\mathbf{H_0}: r_0 = 2$ is generally rejected, at any reasonable level. For fitting GARCH models on these series, the 2QMLE thus seems to outperform the Gaussian QMLE. This could be explained by the fact that the empirical distributions of the residuals $\hat{\eta}_t$ have often larger tails than the Gaussian. Consequently, power Gamma distributions $h_r(x)$ with small values of r are more appropriate than the Gaussian for approximating the actual distribution of η_t.

	CAC	DAX	DJA	DJI	DJT	DJU	FTSE	NIKKEI	SMI	SP500
1990 - 1991	0.88 (0.0)	0.46 (0.0)	0.78 (0.0)	0.74 (0.0)	0.75 (0.0)	0.67 (0.0)	1.30 (0.0)	1.10 (0.0)	0.52 (0.0)	1.00 (0.0)
1992 - 1993	1.20 (0.0)	0.61 (0.0)	1.30 (0.0)	1.30 (0.0)	1.40 (0.0)	1.40 (0.0)	0.88 (0.0)	0.65 (0.0)	0.94 (0.0)	1.50 (0.0)
1994 - 1995	2.10 (0.8)	1.10 (0.0)	1.80 (0.3)	1.30 (0.0)	1.50 (0.0)	1.30 (0.0)	2.20 (0.7)	0.88 (0.0)	1.10 (0.0)	1.40 (0.0)
1996 - 1997	0.51 (0.0)	0.79 (0.0)	0.96 (0.0)	0.91 (0.0)	0.95 (0.0)	0.95 (0.0)	1.30 (0.0)	0.42 (0.0)	0.61 (0.0)	0.82 (0.0)
1998 - 1999	1.60 (0.0)	0.62 (0.0)	1.30 (0.0)	1.30 (0.0)	1.80 (0.5)	NA (NA)	2.00 (0.9)	1.20 (0.0)	0.90 (0.0)	1.20 (0.0)
2000 - 2001	1.10 (0.0)	0.73 (0.0)	0.80 (0.0)	0.91 (0.0)	0.71 (0.0)	1.60 (0.0)	0.61 (0.0)	0.99 (0.0)	0.50 (0.0)	1.20 (0.0)
2002 - 2003	1.50 (0.0)	2.20 (0.9)	2.30 (0.9)	2.00 (1.0)	1.50 (0.0)	1.10 (0.0)	1.30 (0.0)	1.60 (0.0)	1.70 (0.3)	2.00 (1.0)
2004 - 2005	1.20 (0.0)	1.60 (0.1)	NA (NA)	NA (NA)	2.00 (0.1)	1.50 (0.0)	1.20 (0.0)	0.91 (0.0)	0.61 (0.0)	NA (NA)
2006 - 2007	1.60 (0.1)	1.80 (0.4)	0.85 (0.0)	0.72 (0.0)	1.70 (0.2)	0.51 (0.0)	1.00 (0.0)	1.20 (0.0)	1.20 (0.0)	0.52 (0.0)
2008 - 2009	1.20 (0.0)	1.10 (0.0)	0.79 (0.0)	0.81 (0.0)	1.10 (0.0)	1.20 (0.0)	1.30 (0.0)	0.69 (0.0)	0.81 (0.0)	1.10 (0.0)
1990 - 2009	**1.06 (0.0)**	**0.94 (0.0)**	**1.05 (0.0)**	**1.01 (0.0)**	**0.93 (0.0)**	**1.14 (0.0)**	**1.23 (0.0)**	**1.18 (0.0)**	**0.92 (0.0)**	**1.03 (0.0)**

Table 2.6: Estimated variance minimizer r_n and, in parenthesis, the p-value of the test $\mathbf{H_0}: \ r_0 = 2$

2.6 Conclusion

In this paper we investigated the asymptotic properties of a two-stage QMLE, depending on a parameter r, which coincides with the standard QMLE when $r = 2$, and is asymptotically more efficient than the standard QMLE when the optimal value of r is not equal to 2. The CAN of this estimator is obtained under essentially the assumptions required for the CAN of the standard QMLE. We obtained a consistent estimator r_n of the optimal value of r and derived the asymptotic distribution of r_n. A test on the optimal value of r is deduced, which can be used to assess whether, within this class of estimators, the Gaussian QMLE is optimal. It is seen that for most of the daily stock market indices, the two-stage QMLE is more efficient than the standard QMLE.

Several extensions could be considered for future researches. In particular, it could be interesting to define and study a 2QMLE for estimating a wider class of GARCH-type models, in particular those allowing for asymmetries. Another extension would be to propose a 2QMLE for estimating GARCH models identified by the moment condition $E|\eta_t|^r = 1$ with $r < 2$, which would be particularly relevant when η_t does not possess fourth-order moments.

2.7 Proofs

Let $K > 0$ and $\rho \in (0, 1)$ denote constants whose values can change throughout the proofs.

2.7.1 CAN of the generalized QMLE of $\theta_0^{(r)}$

The following result was obtained, under slightly different assumptions, by Berkes and Horváth (2004).

Lemma 2.7.1. *Under Assumptions* **A1-A4**, *we have almost surely*

$$\widehat{\theta}_n^{(r)} \underset{n\to+\infty}{\longrightarrow} \theta_0^{(r)}.$$

If, in addition **A5-A6** *hold, we have:*

$$\sqrt{n}\left(\widehat{\theta}_n^{(r)} - \theta_0^{(r)}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{2}{r}(\mu_{2r}^{(r)} - 1)\{J^{(r)}\}^{-1}\right)$$
$$J^{(r)} = E\left(\frac{r}{2}\phi_t^{(r)}\phi_t^{(r)\prime}\right), \quad \phi_t^{(r)} = \phi_t(\theta_0^{(r)}).$$

By the arguments of the proof of Theorem 2.1 in Francq and Zakoïan (2004), Lemma 2.7.1 follows from the next three lemmas, whose proofs are available from the authors.

Write any $\theta \in \Theta^{(r)}$ in the form $\theta = (\theta_1, \dots, \theta_{p+q+1})'$. We also introduce the notation

$$l_t(\theta) = \log \sigma_t^2(\theta) + \frac{2}{r}\frac{|\epsilon_t|^r}{\sigma_t^r(\theta)}.$$

Lemma 2.7.2. *Under the assumptions of Lemma 2.7.1, for all positive real d, all positive integer k and all indices $i_1, \dots, i_k \in \{1, \dots, p+q+1\}$, we have*

$$E \sup_{\theta\in\Theta^*}\left|\frac{1}{\sigma_t^2(\theta)}\frac{\partial^k \sigma_t^2(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}}\right|^d < \infty, \quad E \sup_{\theta\in\mathcal{V}(\theta_0^{(r)})}\left|\frac{\sigma_t^2(\theta_0^{(r)})}{\sigma_t^2(\theta)}\right|^d < \infty, \tag{2.7.1}$$

for any compact subset Θ^* of the interior of $\Theta^{(r)}$ such that $\theta_0^{(r)} \in \Theta^*$, and for some neighborhood $\mathcal{V}(\theta_0^{(r)})$ of $\theta_0^{(r)}$, and

$$E\left\|\frac{\partial l_t}{\partial\theta}\frac{\partial l_t}{\partial\theta'}(\theta_0^{(r)})\right\| < \infty, \qquad E\left\|\frac{\partial^2 l_t}{\partial\theta\partial\theta'}(\theta_0^{(r)})\right\| < \infty \tag{2.7.2}$$

Lemma 2.7.3. *Under the assumptions of Lemma 2.7.1, for all positive real d, all positive integer k and all indices $i_1, \dots, i_k \in \{1, \dots, p+q+1\}$, we have*

$$\sup_{\theta\in\Theta^{(r)}}\left|\sigma_t^2(\theta) - \tilde{\sigma}_t^2(\theta)\right| \le K\rho^t \quad a.s., \qquad \sup_{\theta\in\Theta^{(r)}}\left|\frac{\partial^k \sigma_t^2(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}} - \frac{\partial^k \tilde{\sigma}_t^2(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}}\right| \le K\rho^t \quad a.s., \tag{2.7.3}$$

$$\sup_{\theta\in\Theta^{(r)}}\left|\frac{\sigma_t^d(\theta)}{\tilde{\sigma}_t^d(\theta)} - 1\right| \le K\rho^t \quad a.s., \tag{2.7.4}$$

$$\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\frac{\partial \tilde{l}_t(\theta_0^{(r)})}{\partial\theta} - \frac{\partial l_t(\theta_0^{(r)})}{\partial\theta}\right\}\right\| \to 0 \quad in\ probability, \tag{2.7.5}$$

and, for some neighborhood $\mathcal{V}(\theta_0^{(r)})$ of $\theta_0^{(r)}$,

$$\sup_{\theta\in\mathcal{V}(\theta_0^{(r)})}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right\}\right\|\to 0 \quad a.s.$$

Lemma 2.7.4. *Under the assumptions of Lemma 2.7.1, there exists a neighborhood $\mathcal{V}(\theta_0^{(r)})$ of $\theta_0^{(r)}$ such that, for $i,j,k\in\{1,\cdots,p+q+1\}$,*

$$E\sup_{\theta\in\mathcal{V}(\theta_0^{(r)})}\left|\frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|<+\infty.$$

2.7.2 Proof of Theorem 2.2.1

The following lemma gives the joint distribution of the QML estimator of $\theta_0^{(r)}$ and the empirical mean of the squared standardized residuals.

Lemma 2.7.5. *Under Assumptions **A1-A6**, we have*

$$\begin{pmatrix}\sqrt{n}\left(\widehat{\theta}_n^{(r)}-\theta_0^{(r)}\right)\\ \sqrt{n}\left(\hat{\mu}_{2,n}^{(r)}-\mu_2^{(r)}\right)\end{pmatrix}\xrightarrow{\mathcal{L}}\mathcal{N}\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\frac{2}{r}\left(\mu_{2r}^{(r)}-1\right)\{J^{(r)}\}^{-1} & \Gamma_r'\\ \Gamma_r & \Lambda_r\end{pmatrix}\right\}, \tag{2.7.6}$$

with

$$\Gamma_r=\frac{2}{r}b_r\overline{\theta}_0^{(r)},\quad \Lambda_r=\mu_4^{(r)}-\mu_2^{(r)2}-\frac{2}{r}\mu_2^{(r)}\left(b_r+\mu_{2+r}^{(r)}-\mu_2^{(r)}\right),$$
$$b_r=\mu_{2+r}^{(r)}-\mu_2^{(r)}-\frac{2}{r}\mu_2^{(r)}(\mu_{2r}^{(r)}-1),$$

and $\overline{\theta}_0^{(r)}=(\omega_0^{(r)},\alpha_{01}^{(r)},\ldots,\alpha_{0q}^{(r)},0,\ldots,0)'$.

Proof. We have

$$\sqrt{n}\left(\hat{\mu}_{2,n}^{(r)}-\mu_2^{(r)}\right)=n^{-\frac{1}{2}}\sum_{t=1}^{n}\left(\frac{\epsilon_t^2}{\sigma_t^2(\widehat{\theta}_n^{(r)})}-\mu_2^{(r)}\right)+n^{-\frac{1}{2}}\sum_{t=1}^{n}\epsilon_t^2\left(\frac{1}{\tilde{\sigma}_t^2(\widehat{\theta}_n^{(r)})}-\frac{1}{\sigma_t^2(\widehat{\theta}_n^{(r)})}\right).$$

Using the first inequality in (2.7.3), it is easily seen that

$$n^{-\frac{1}{2}}\sum_{t=1}^{n}\epsilon_t^2\left(\frac{1}{\tilde{\sigma}_t^2(\widehat{\theta}_n^{(r)})}-\frac{1}{\sigma_t^2(\widehat{\theta}_n^{(r)})}\right)=o_P(1).$$

This result and a Taylor expansion show that, for some $\overline{\theta}$ between $\widehat{\theta}_n^{(r)}$ and $\theta_0^{(r)}$,

$$\sqrt{n}\left(\hat{\mu}_{2,n}^{(r)}-\mu_2^{(r)}\right)=\sqrt{n}\left(\mu_{2,n}^{(r)}-\mu_2^{(r)}\right)-\left\{n^{-1}\sum_{t=1}^{n}\epsilon_t^2\left(\frac{1}{\sigma_t^4(\overline{\theta})}\frac{\partial\sigma_t^2(\overline{\theta})}{\partial\theta'}\right)\right\}\sqrt{n}(\widehat{\theta}_n^{(r)}-\theta_0^{(r)})+o_P(1), \tag{2.7.7}$$

where $\mu_{2,n}^{(r)}=n^{-1}\sum_{t=1}^{n}|\eta_t^{(r)}|^2$. We will now show that

$$n^{-1}\sum_{t=1}^{n}\epsilon_t^2\left(\frac{1}{\sigma_t^4(\overline{\theta})}\frac{\partial\sigma_t^2(\overline{\theta})}{\partial\theta}\right)\to\mu_2^{(r)}\phi^{(r)\prime}\quad a.s., \tag{2.7.8}$$

where $\phi^{(r)}=E\phi_t^{(r)}$. By another Taylor expansion, for any $i\in\{1,\ldots,p+q+1\}$ there exists θ^* between $\widehat{\theta}_n^{(r)}$ and $\theta_0^{(r)}$ such that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2\left(\frac{1}{\sigma_t^4(\overline{\theta})}\frac{\partial\sigma_t^2(\overline{\theta})}{\partial\theta_i}\right) &= \frac{1}{n}\sum_{t=1}^{n}\frac{\epsilon_t^2}{\sigma_t^2(\theta_0^{(r)})}\frac{1}{\sigma_t^2(\theta_0^{(r)})}\frac{\partial\sigma_t^2(\theta_0^{(r)})}{\partial\theta_i}\\ &\quad+\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2\frac{\partial}{\partial\theta'}\left(\frac{1}{\sigma_t^4(\theta^*)}\frac{\partial\sigma_t^2(\theta^*)}{\partial\theta_i}\right)(\overline{\theta}-\theta_0^{(r)}).\end{aligned}$$

Since $\frac{\epsilon_t^2}{\sigma_t^2(\theta_0^{(r)})}=\eta_t^{(r)2}$, using the independence between $\eta_t^{(r)}$ and $\sigma_t^{(r)}$ and the ergodic theorem, we obtain the a.s.convergence of the first term to $\mu_2^{(r)}\phi^{(r)}$. In view of (2.7.1), we have

$$E\sup_{\theta\in\mathcal{V}(\theta_0^{(r)})}\left\|\epsilon_t^2\frac{\partial}{\partial\theta}\left(\frac{1}{\sigma_t^4(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\right)\right\|<\infty.$$

We conclude that (2.7.8) holds, using the a.s. convergence of $\left\|\overline{\theta}-\theta_0^{(r)}\right\|$ to 0 and the ergodic theorem.

It follows from (2.7.7)-(2.7.8) that

$$\sqrt{n}\left(\hat{\mu}_{2,n}^{(r)}-\mu_2^{(r)}\right)=\sqrt{n}\left(\mu_{2,n}^{(r)}-\mu_2^{(r)}\right)-\mu_2^{(r)}\phi^{(r)\prime}\sqrt{n}\left(\widehat{\theta}_n^{(r)}-\theta_0^{(r)}\right)+o_P(1).$$

Now, by standard arguments we have

$$\sqrt{n}\left(\widehat{\theta}_n^{(r)}-\theta_0^{(r)}\right)=\{J^{(r)}\}^{-1}n^{-1/2}\sum_{t=1}^{n}\left(|\eta_t^{(r)}|^r-1\right)\phi_t^{(r)}+o_P(1).$$

Moreover,

$$\sqrt{n}\left(\mu_{2,n}^{(r)} - \mu_2^{(r)}\right) = n^{-1/2}\sum_{t=1}^{n}\left(|\eta_t^{(r)}|^2 - \mu_2^{(r)}\right).$$

Then

$$\begin{aligned}\sqrt{n}\begin{pmatrix}\hat{\mu}_{2,n}^{(r)} - \mu_2^{(r)} \\ \widehat{\theta}_n^{(r)} - \theta_0^{(r)}\end{pmatrix} &= \begin{pmatrix}1 & -\mu_2^{(r)}\phi^{(r)\prime} \\ 0 & I_{p+q+1}\end{pmatrix}\sqrt{n}\begin{pmatrix}\mu_{2,n}^{(r)} - \mu_2^{(r)} \\ \widehat{\theta}_n^{(r)} - \theta_0^{(r)}\end{pmatrix} + o_P(1) \\ &= \begin{pmatrix}1 & -\mu_2^{(r)}\phi^{(r)\prime}\{J^{(r)}\}^{-1} \\ 0 & \{J^{(r)}\}^{-1}\end{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}|\eta_t^{(r)}|^2 - \mu_2^{(r)} \\ \phi_t^{(r)}\left(|\eta_t^{(r)}|^r - 1\right)\end{pmatrix} + o_P(1).\end{aligned}$$

By the already used CLT,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}|\eta_t^{(r)}|^2 - \mu_2^{(r)} \\ \phi_t^{(r)}\left(|\eta_t^{(r)}|^r - 1\right)\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\mu_4^{(r)} - \mu_2^{(r)2} & \phi^{(r)\prime}\left(\mu_{r+2}^{(r)} - \mu_2^{(r)}\right) \\ \left(\mu_{r+2}^{(r)} - \mu_2^{(r)}\right)\phi^{(r)} & \frac{2}{r}\left(\mu_{2r}^{(r)} - 1\right)J^{(r)}\end{pmatrix}\right\}.$$

The asymptotic normality (2.7.6) follows with

$$\Gamma_r = b_r\phi^{(r)\prime}\{J^{(r)}\}^{-1} \quad \text{and} \quad \Lambda_r = \mu_4^{(r)} - \mu_2^{(r)2} - \mu_2^{(r)}\left(b_r + \mu_{2+r}^{(r)} - \mu_2^{(r)}\right)\phi^{(r)\prime}\{J^{(r)}\}^{-1}\phi^{(r)}.$$

Now, denoting by θ_1 the sub-vector of the first $q+1$ components of θ, note that

$$\theta_1'\frac{\partial\sigma_t^2(\theta)}{\partial\theta_1} = \omega + \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2(\theta) + \sum_{j=1}^{p}\beta_j\theta_1'\frac{\partial\sigma_{t-j}^2(\theta)}{\partial\theta_1} = \sigma_t^2(\theta)$$

using Assumption **A2**. This relation can be written as

$$\overline{\theta}'\frac{\partial\sigma_t^2(\theta)}{\partial\theta} = \sigma_t^2(\theta) \quad \text{with} \quad \overline{\theta}' = (\theta_1', 0_p')$$

or equivalently $\overline{\theta}'\phi_t(\theta) = 1$ a.s. We finally obtain

$$\overline{\theta}_0^{(r)\prime}J^{(r)} = \frac{r}{2}E\overline{\theta}_0^{(r)\prime}\phi_t^{(r)}\phi_t^{(r)\prime} = \frac{r}{2}\phi^{(r)\prime}, \quad \phi^{(r)\prime}\{J^{(r)}\}^{-1} = \frac{2}{r}\overline{\theta}_0^{(r)\prime}, \quad \phi^{(r)\prime}\{J^{(r)}\}^{-1}\phi^{(r)} = \frac{2}{r},$$

which gives the expression of Γ_r and Λ_r and completes the proof. □

Proof of Theorem 2.2.1. The proof of the a.s. convergence relies on arguments already used and is available from the authors. The asymptotic normality of $\hat{\theta}_{n,r}$ follows from (2.1.1), (2.2.3), (2.7.6), and the a.s. convergence of $\widehat{B}_n^{(r)}$ to $B^{(r)}$. Tedious computations available from the authors yield the announced formula for Σ_r. □

2.7.3 Proof of Corollary 2.2.1

The corollary is a direct consequence of Theorem 2.2.1 and of (4.9) in Francq and Zakoian (2010), where it is shown that the matrix $J^{-1} - \overline{\theta}_0\overline{\theta}_0'$ is semi-definite positive.

2.7.4 Proof of Theorem 2.3.1

For $u > 0$ and $v \geq 0$, we use the convention $0^u(\log|0|)^v = 0$. Let

$$m(u,v) = E\left\{|\eta_t|^u\left(\log|\eta_t|\right)^v\right\}, \tag{2.7.9}$$

when the expectation exists, and let

$$\widehat{m}_n(u,v) = \frac{1}{n}\sum_{t=1}^n |\widehat{\eta}_t|^u\left(\log|\widehat{\eta}_t|\right)^v \quad \text{with} \quad \widehat{\eta}_t = \frac{\epsilon_t}{\tilde{\sigma}_t\left(\widehat{\theta}_{n,2}\right)}.$$

We now need to introduce the notations

$$\tilde{m}_n(u,v,\theta) = \frac{1}{n}\sum_{t=1}^n\left|\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}\right|^u\left(\log\left|\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}\right|\right)^v, \quad m_n(u,v,\theta) = \frac{1}{n}\sum_{t=1}^n\left|\frac{\epsilon_t}{\sigma_t(\theta)}\right|^u\left(\log\left|\frac{\epsilon_t}{\sigma_t(\theta)}\right|\right)^v,$$
$$m_\infty(u,v,\theta) = E\left|\frac{\epsilon_t}{\sigma_t(\theta)}\right|^u\left(\log\left|\frac{\epsilon_t}{\sigma_t(\theta)}\right|\right)^v, \qquad \tilde{g}_n(r,\theta) = \left(\frac{2}{r}\right)^2\left(\frac{\tilde{m}_n(2r,0,\theta)}{\tilde{m}_n(r,0,\theta)^2} - 1\right),$$

and let $g_n(r,\theta)$ (resp. $g_\infty(r,\theta)$) be obtained by replacing $\tilde{m}_n(u,v,\theta)$ by $m_n(u,v,\theta)$ (resp. by $m_\infty(u,v,\theta)$) in $\tilde{g}_n(r,\theta)$. We thus have $\hat{m}_n(u,v) = \tilde{m}_n(u,v,\widehat{\theta}_{n,2})$, $\hat{g}_n(r) = \tilde{g}_n(r,\widehat{\theta}_{n,2})$, and $m(u,v) = m_\infty(u,v,\theta_0)$. Before establishing the consistency of the estimator r_n, we state several lemmas.

Lemma 2.7.6. *Let $R^* = [\underline{r}, 2\overline{r}]$. Under the assumptions of Theorem 2.3.1 we have, for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0, almost surely*

$$\overline{\lim_n} \sup_{s\in R^*,\ \theta\in\mathcal{V}(\theta_0)} |\tilde{m}_n(s,0,\theta)| < +\infty, \qquad \overline{\lim_n} \sup_{s\in R^*,\ \theta\in\mathcal{V}(\theta_0)} \left\|\frac{\partial\tilde{m}_n}{\partial\theta}(s,0,\theta)\right\| < +\infty, \tag{2.7.10}$$

$$\overline{\lim_n} \sup_{s\in R^*,\ \theta\in\mathcal{V}(\theta_0)} \left\|\frac{\partial\tilde{m}_n}{\partial s}(s,0,\theta)\right\| < +\infty. \tag{2.7.11}$$

Proof. We have, by Lemma 2.7.3,

$$\sup_{\theta\in\mathcal{V}(\theta_0),\ s\in R^*}\frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)} \leq K \sup_{\theta\in\mathcal{V}(\theta_0),\ s\in R^*}\left(\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right).$$

Now

$$\begin{aligned} E\left[\sup_{\theta\in\mathcal{V}(\theta_0),\ s\in R^*}\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right] &= E\left[\sup_{\theta\in\mathcal{V}(\theta_0),\ s\in R^*}\left\{\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\mathbb{1}_{|\epsilon_t|^s\leq\sigma_t^s(\theta)} + \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\mathbb{1}_{|\epsilon_t|^s>\sigma_t^s(\theta)}\right\}\right] \\ &\leq 1 + E\left[\sup_{\theta\in\mathcal{V}(\theta_0)}\frac{|\epsilon_t|^{2\overline{r}}}{\sigma_t^{2\overline{r}}(\theta)}\right] < \infty. \end{aligned}$$

The last inequality follows from the definition of $\overline{r}$ and the second inequality in (2.7.1). It follows that, by the ergodic theorem

$$\overline{\lim_n} \sup_{\theta\in\mathcal{V}(\theta_0),\ s\in R^*}\frac{1}{n}\sum_{t=1}^n\frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)} \leq E\left[\sup_{\theta\in\mathcal{V}(\theta_0),\ s\in R^*}\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right] < +\infty,$$

which proves the first inequality in (2.7.10).

For d large enough, we have $E|\eta_t|^{s\frac{d+1}{d}} < +\infty$ for all $s \in R^*$. Hence, by already used arguments,

$$\left\|\sup_{\theta\in\mathcal{V}(\theta_0),\ s\in R^*}\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}\right\|_{\frac{d+1}{d}} < \infty. \tag{2.7.12}$$

We have

$$\frac{\partial\tilde{m}_n}{\partial\theta}(s,0,\theta) = -\frac{s}{2}\frac{1}{n}\sum_{t=1}^n\tilde{\phi}_t(\theta)\frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)}.$$

Now,

$$\left\|\tilde{\phi}_t(\theta)\right\| \leq \frac{1}{\tilde{\sigma}_t^2(\theta)}\left\|\frac{\partial\tilde{\sigma}_t^2}{\partial\theta}(\theta) - \frac{\partial\sigma_t^2}{\partial\theta}(\theta)\right\| + \|\phi_t(\theta)\|\frac{\sigma_t^2(\theta)}{\tilde{\sigma}_t^2(\theta)} \leq \|\phi_t(\theta)\|(1+K\rho^t) + K\rho^t.$$

Therefore,

$$\left\|\frac{\partial\tilde{m}_n}{\partial\theta}(s,0,\theta)\right\| \leq \frac{s}{2}\frac{1}{n}\sum_{t=1}^n\left(\|\phi_t(\theta)\| + K\rho^t\right)\frac{|\epsilon_t|^s}{\sigma_t^s(\theta)}(1+K\rho^t). \tag{2.7.13}$$

From the first inequality in (2.7.1), we have for any integer $d > 0$,

$$\forall k \in \{1, \dots, p+q+1\}, \quad E \sup_{\theta \in \mathcal{V}(\theta_0)} |\phi_{t,k}|^d < \infty,$$

where $\phi_{t,k}(\theta)$ is the k^{th} component of the vector $\phi_t(\theta)$.

Then, by Hölder inequality and (2.7.12), we obtain

$$E \sup_{\theta \in V(\theta_0),\ s \in R^*} \left| \phi_{t,k}(\theta) \frac{|\epsilon_t|^s}{\sigma_t^s(\theta)} \right| < \infty.$$

Thus, by the ergodic theorem and (2.7.13), we obtain the second part of (2.7.10). Finally, using

$$\frac{\partial \tilde{m}_n}{\partial s}(s, 0, \theta) = \frac{1}{n} \sum_{t=1}^{n} \log \left(\frac{|\epsilon_t|}{\tilde{\sigma}_t(\theta)} \right) \frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)} = \tilde{m}_n(s, 1, \theta),$$

(2.7.11) can be proven by similar arguments. □

Lemma 2.7.7. *Under the assumptions of Theorem 2.3.1 we have, for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0, almost surely*

$$\overline{\lim_n} \sup_{s \in R,\ \theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial \tilde{g}_n}{\partial \theta}(s, \theta) \right\| < +\infty, \qquad \overline{\lim_n} \sup_{s \in R,\ \theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial \tilde{g}_n}{\partial r}(s, \theta) \right\| < +\infty.$$

Proof. We have

$$\frac{\partial \tilde{g}_n}{\partial \theta}(s, \theta) = \frac{4}{s^2} \frac{\frac{\partial \tilde{m}_n}{\partial \theta}(2s, 0, \theta) \tilde{m}_n(s, 0, \theta) - \tilde{m}_n(2s, 0, \theta) \frac{\partial \tilde{m}_n}{\partial \theta}(s, 0, \theta)}{\tilde{m}_n^3(s, 0, \theta)}.$$

and

$$\frac{\partial \tilde{g}_n}{\partial r}(s, \theta) = -\frac{8}{s^3} \left(\frac{\tilde{m}_n(2s, 0, \theta)}{\tilde{m}_n^2(s, 0, \theta)} - 1 \right) + \frac{8}{s^2} \frac{\tilde{m}_n(2s, 1, \theta) \tilde{m}_n(s, 0, \theta) - \tilde{m}_n(s, 1, \theta) \tilde{m}_n(2s, 0, \theta)}{\tilde{m}_n^3(s, 0, \theta)}. \tag{2.7.14}$$

In view of Lemma 2.7.6, it suffices to note that, for n large enough,

$$\inf_{s \in R,\ \theta \in \mathcal{V}(\theta_0)} \tilde{m}_n(s, 0, \theta) > K. \tag{2.7.15}$$

A Taylor expansion yields

$$\tilde{m}_n(s,0,\theta) = \frac{1}{n}\sum_{t=1}^{n}\frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta)} = \frac{1}{n}\sum_{t=1}^{n}\frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta_0)} + (\theta-\theta_0)'\frac{\partial \tilde{m}_n(s,0,\theta^*)}{\partial\theta},$$

where θ^* is between θ and θ_0. Then using Lemma 2.7.3 and the law of large numbers, we obtain the almost sure convergence of the first term to $m(s,0) > 0$. For a large enough n, we have $\frac{1}{n}\sum_{t=1}^{n}\frac{|\epsilon_t|^s}{\tilde{\sigma}_t^s(\theta_0)} > K$ with $K > 0$. Furthermore using equation (2.7.10) of Lemma 2.7.6 and using a small enough neighborhood of θ_0, we obtain the result (2.7.15). □

Proof of the consistency of r_n. Let $r \neq r_0$ and let $V_k(r) = (r - 1/k, r + 1/k)$ for any positive integer k. For $s \in V_k(r)$, $\tilde{r}$ between s and r, and θ between $\widehat{\theta}_{n,2}$ and θ_0,

$$\hat{g}_n(s) = \tilde{g}_n\left(s, \widehat{\theta}_{n,2}\right) = \tilde{g}_n(r,\theta_0) + \frac{\partial \tilde{g}_n}{\partial r}(\tilde{r},\theta)(s-r) + \frac{\partial \tilde{g}_n}{\partial \theta'}(\tilde{r},\theta)(\widehat{\theta}_{n,2}-\theta_0),$$

and thus

$$\begin{aligned}\underline{\lim}_n \inf_{s\in V_k(r)} \hat{g}_n(s) \;\geq\; & \underline{\lim}_n\, \tilde{g}_n(r,\theta_0) - \overline{\lim}_n \sup_{s\in V_k(r),\ \theta\in\mathcal{V}(\theta_0)} \left\|\frac{\partial \tilde{g}_n}{\partial\theta}(s,\theta)\right\| \|\widehat{\theta}_{n,2}-\theta_0\| \\ & - \overline{\lim}_n \sup_{s\in V_k(r),\ \theta\in\mathcal{V}(\theta_0)} \left|\frac{\partial \tilde{g}_n}{\partial r}(s,\theta)\right| |s-r|.\end{aligned}$$

By already used argument concerning the initial values, we have $\tilde{g}_n(r,\theta_0) = g_n(r,\theta_0) + o(1)$ a.s. The law of large numbers shows that $g_n(r,\theta_0) \to g(r)$ a.s. We thus have $\tilde{g}_n(r,\theta_0) = g(r) + o(1)$ a.s. Now, if we choose a large enough k, using **A7**, Lemma 2.7.7 and the consistency of $\widehat{\theta}_{n,2}$ to θ_0, we obtain for sufficiently small $\varepsilon > 0$

$$\underline{\lim}_n \inf_{s\in V_k(r)} \hat{g}_n(s) \geq g(r) - \varepsilon > g(r_0) \quad \text{a.s.}$$

We can now conclude the consistency proof with a standard compactness argument. □

Now, we turn to the asymptotic normality of r_n. A Taylor expansion gives

$$\frac{\partial \hat{g}_n}{\partial r}(r_n) = 0 = \frac{\partial \tilde{g}_n}{\partial r}(r_0,\theta_0) + \frac{\partial^2 \tilde{g}_n}{\partial r\partial\theta'}(\tilde{r},\overline{\theta})\left(\widehat{\theta}_{n,2}-\theta_0\right) + \frac{\partial^2 \tilde{g}_n}{\partial r^2}(\tilde{r},\overline{\theta})(r_n - r_0), \tag{2.7.16}$$

with $\overline{\theta}$ between θ_0 and $\widehat{\theta}_{n,2}$ and $\tilde{r}$ between r_n and r_0.

We will now prove the following lemmas.

Lemma 2.7.8. *Under the assumptions of Theorem 2.3.1, for some positive number* ξ_{r_0}*,*

$$\sqrt{n}\frac{\partial \tilde{g}_n}{\partial r}(r_0, \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \xi_{r_0}).$$

Proof. By already used arguments, we have

$$\frac{\partial \tilde{g}_n}{\partial r}(r_0, \theta_0) - \frac{\partial g_n}{\partial r}(r_0, \theta_0) = o_P(n^{-1/2}).$$

Therefore, we replace $\tilde{g}_n$ by g_n in the rest of the proof. Similarly to (2.7.14), we have

$$\begin{aligned}\frac{\partial g_n}{\partial r}(r_0, \theta_0) &= -\frac{8}{r_0^3}\left(\frac{m_n(2r_0, 0, \theta_0)}{m_n^2(r_0, 0, \theta_0)} - 1\right) \\ &\quad + \frac{8}{r_0^2}\frac{m_n(2r_0, 1, \theta_0)m_n(r_0, 0, \theta_0) - m_n(r_0, 1, \theta_0)m_n(2r_0, 0, \theta_0)}{m_n^3(r_0, 0, \theta_0)}.\end{aligned}$$

For $u, u' > 0$ and $v, v' \geq 0$ we have

$$\mathrm{Cov}(m_n(u, v, \theta_0), m_n(u', v', \theta_0)) = \frac{C(u, v, u', v')}{n},$$

where $C(u, v, u', v') = m(u + u', v + v') - m(u, v)m(u', v')$. The standard CLT entails

$$\begin{pmatrix} \sqrt{n}(m_n(r_0, 0, \theta_0) - m(r_0, 0)) \\ \sqrt{n}(m_n(2r_0, 0, \theta_0) - m(2r_0, 0)) \\ \sqrt{n}(m_n(r_0, 1, \theta_0) - m(r_0, 1)) \\ \sqrt{n}(m_n(2r_0, 1, \theta_0) - m(2r_0, 1)) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0_4, \Omega_{r_0}\right)$$

with

$$\Omega_{r_0} = \begin{pmatrix} C(r_0, 0, r_0, 0) & C(r_0, 0, 2r_0, 0) & C(r_0, 0, r_0, 1) & C(r_0, 0, 2r_0, 1) \\ C(r_0, 0, 2r_0, 0) & C(2r_0, 0, 2r_0, 0) & C(2r_0, 0, r_0, 1) & C(2r_0, 0, 2r_0, 1) \\ C(r_0, 0, r_0, 1) & C(2r_0, 0, r_0, 1) & C(r_0, 1, r_0, 1) & C(r_0, 1, 2r_0, 1) \\ C(r_0, 0, 2r_0, 1) & C(2r_0, 0, 2r_0, 1) & C(r_0, 1, 2r_0, 1) & C(2r_0, 1, 2r_0, 1) \end{pmatrix}$$

Then, by the delta method, we obtain

$$\xi_{r_0} = \Delta'_{r_0}\Omega_{r_0}\Delta_{r_0}, \tag{2.7.17}$$

where, after a simplification following from $\frac{\partial g}{\partial r}(r_0) = 0$,

$$\begin{aligned}\Delta_{r_0} \quad = \quad & \left(\frac{8\left(r_0 m(2r_0,1) + 3m(r_0,0)^2 - m(2r_0,0)\right)}{r_0^3 m(r_0,0)^3}, \; -\frac{8\left(m(r_0,0) + r_0 m(r_0,1)\right)}{r_0^3 m^3(r_0,0)},\right. \\ & \left. -\frac{8m(2r_0,0)}{r_0^2 m^3(r_0,0)}, \; \frac{8}{r_0^2 m^2(r_0,0)}\right)'\end{aligned}$$

It remains to show that Ω_r, the covariance matrix of the vector

$$Z = \left(|\eta_t|^r, \; |\eta_t|^{2r}, \; |\eta_t|^r \log|\eta_t|, \; |\eta_t|^{2r}\log|\eta_t|\right)',$$

is positive definite. We argue by contradiction and assume that Ω_r is singular. Then there exist a vector $c = (c_1, \ldots, c_4)' \neq 0$ and a constant c_5 such that $c'Z = c_5$ a.s. Thus $g(x) := c_1 x + c_2 x^2 + (c_3/r)x\log x + (c_4/r)x^2\log x - c_5 = 0$ for $P_{|\eta|^r}$-almost all x. The third-order derivative of g is null for at most one value of x. It follows that g is null for at most 4 values of x, which is in contradiction with Assumption **A3**. $\square$

Lemma 2.7.9. *Under the assumptions of Theorem 2.3.1,*

$$\frac{\partial^2 \tilde{g}_n}{\partial r^2}(\tilde{r}, \overline{\theta}) \overset{a.s.}{\to} \frac{\partial^2 g_\infty}{\partial r^2}(r_0, \theta_0).$$

Proof. As in the consistency proof, by a Taylor expansion we have

$$\frac{\partial^2 \tilde{g}_n}{\partial r^2}(\tilde{r}, \overline{\theta}) = \frac{\partial^2 \tilde{g}_n}{\partial r^2}(r_0, \theta_0) + \frac{\partial^3 \tilde{g}_n}{\partial r^3}(s, \theta)(\tilde{r} - r_0) + \frac{\partial^3 \tilde{g}_n}{\partial r^2 \partial \theta}(s, \theta)(\overline{\theta} - \theta_0),$$

with s between $\tilde{r}$ and r_0 and θ between $\overline{\theta}$ and θ_0. We obtain, by already used arguments, for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0, and for some large enough k,

$$\overline{\lim_n} \sup_{s\in V_k(r_0),\ \theta\in\mathcal{V}(\theta_0)} \left\| \frac{\partial^3 \tilde{g}_n}{\partial r^2 \partial\theta}(s,\theta) \right\| < +\infty, \tag{2.7.18}$$

$$\overline{\lim_n} \sup_{s\in V_k(r),\ \theta\in\mathcal{V}(\theta_0)} \left\| \frac{\partial^3 \tilde{g}_n}{\partial r^3}(s,\theta) \right\| < +\infty. \tag{2.7.19}$$

It follows from the convergence of $(r_n, \widehat{\theta}_{n,2})$ to (r_0, θ_0) that

$$\frac{\partial^2 \tilde{g}_n}{\partial r^2}(\tilde{r}, \overline{\theta}) = \frac{\partial^2 \tilde{g}_n}{\partial r^2}(r_0, \theta_0) + o(1), \quad a.s.$$

The almost sure convergence of $\frac{\partial^2 \tilde{g}_n}{\partial r^2}(r_0, \theta_0)$ to $\frac{\partial^2 g_\infty}{\partial r^2}(r_0, \theta_0)$ is obtained by the law of large numbers, which completes the proof. □

Now we give an explicit expression to the second-order derivative of g_∞. Note that the first-order derivative is obtained from (2.7.14) by replacing the $\tilde{m}_n(u, v, \theta_0)$ by $m(u, v)$. By another differentiation we obtain, using the elementary equality $\partial m(u, v, \theta)/\partial s = m(u, v+1, \theta)$,

$$\begin{aligned}
\frac{\partial^2 g_\infty}{\partial r^2}(r_0, \theta_0) \quad = \quad & \frac{24}{r_0^4}\left(\frac{m(2r_0, 0)}{m^2(r_0, 0)} - 1\right) + \frac{8}{r_0^2}\frac{1}{m^3(r_0, 0)}\Big\{2m(2r_0, 2)m(r_0, 0) \\
& -m(2r_0, 1)m(r_0, 1) - m(2r_0, 0)m(r_0, 2)\Big\} \\
& -\frac{32}{r_0^3}\frac{m(2r_0, 1)m(r_0, 0) - m(2r_0, 0)m(r_0, 1)}{m(r_0, 0)^3} \\
& -\frac{24}{r_0^2}\frac{m(r_0, 1)}{m^4(r_0, 0)}\left\{m(2r_0, 1)m(r_0, 0) - m(2r_0, 0)m(r_0, 1)\right\}.
\end{aligned}$$

We know that $\frac{\partial g}{\partial r}(r_0) = 0$. This allows us to simplify the previous equation.

$$\begin{aligned}
\frac{\partial^2 g_\infty}{\partial r^2}(r_0, \theta_0) \quad = \quad & \frac{8}{r_0^2}\frac{2m(2r_0, 2)m(r, 0) - m(2r_0, 1)m(r_0, 1) - m(2r_0, 0)m(r_0, 2)}{m^3(r_0, 0)} \\
& -\frac{8}{r_0^3}\left(\frac{m(2r_0, 0)}{m^2(r_0, 0)} - 1\right)\left(\frac{1}{r_0} + 3\frac{m(r_0, 1)}{m(r_0, 0)}\right).
\end{aligned} \tag{2.7.20}$$

Lemma 2.7.10. *Under the assumptions of Theorem 2.3.1,*

$$\frac{\partial^2 \tilde{g}_n}{\partial r \partial\theta}(\tilde{r}, \overline{\theta}) \stackrel{a.s.}{\to} \frac{\partial^2 g_\infty}{\partial r \partial\theta}(r_0, \theta_0) = 0$$

Proof. We write another Taylor expansion, use equations (2.7.18) and (2.7.19) and apply the law of large numbers to show the convergence. Furthermore

$$\begin{aligned}\frac{\partial^2 g_\infty}{\partial r \partial \theta}(r_0,\theta_0) &= \frac{4}{r_0}\frac{1}{m^4(r_0,0)}\Big\{m(r_0,0)\left\{2m_\phi(r_0,0)m(2r_0,1)+m_\phi(r_0,1)m(2r_0,0)\right.\\ &\qquad \left. -m(r_0,1)m_\phi(2r_0,0) -2m_\phi(2r_0,1)m(r_0,0)\right\}\\ &\qquad +3m(r_0,1)\left\{m_\phi(r_0,0)m(2r_0,0)-m(r_0,0)m_\phi(2r_0,0)\right\}\Big\}\\ &\quad -\frac{8}{r_0^2}\frac{m_\phi(r_0,0)m(2r_0,0)-m(r_0,0)m_\phi(2r_0,0)}{m^3(r_0,0)}\end{aligned}$$

with $m_\phi(u,v)=E\left[\phi_t|\eta_t|^u(\log|\eta_t|)^v\right]$. Then with the independence between ϕ_t and η_t, we obtain the result. $\square$

With (2.7.16) and the previous lemmas, the convergence in law (2.3.2) follows with

$$\tau_{r_0}=\xi_{r_0}\left\{\frac{\partial^2 g_\infty}{\partial r^2}(r_0,\theta_0)\right\}^{-2} \tag{2.7.21}$$

where the denominator is displayed in (2.7.20), and the numerator in (2.7.17).

2.7.5 Proof of Lemma 2.3.1

Note that for fixed numbers $x_1,\dots,x_n$ such that $x_{(n)}:=\max_{i=1,\dots,n}|x_i|>0$, we have

$$\lim_{r\to\infty}\frac{|x_1|^r+\cdots+|x_n|^r}{x_{(n)}^r}=k_x$$

where k_x is the number of $|x_i|$ which are equal to $x_{(n)}$. With an obvious notation, we thus have

$$\lim_{r\to\infty}\frac{\hat{m}_n(2r,0)}{\hat{m}_n^2(r,0)}=\frac{n}{k_{\hat{\eta}}},$$

and the result immediately follows.

Bibliography

J.M. Bardet and O. Wintenberger. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics*, 37(5B):2730–2759, 2009.

I. Berkes and L. Horváth. The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics*, 32(2):633–655, 2004.

I. Berkes, L. Horvath, and P. Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, pages 201–227, 2003.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

X. Chen. Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, 6:5549–5632, 2007.

J. Di and A. Gangopadhyay. On the efficiency of a semi-parametric GARCH model. *The Econometrics Journal*, 14(2):257–277, 2011.

F.C. Drost and C.A.J. Klaassen. Efficient estimation in semiparametric GARCH models. *Journal of Econometrics*, 81(1):193–221, 1997.

R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, pages 987–1007, 1982.

R.F. Engle and G. Gonzalez-Rivera. Semiparametric ARCH models. *Journal of Business & Economic Statistics*, 9(4):345–359, 1991.

C. Francq and J.M. Zakoïan. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–638, 2004.

C. Francq and J.M. Zakoian. Optimal predictions of powers of conditionally heteroskedastic processes. MPRA Paper 22155, University Library of Munich, Germany, April 2010. URL http://ideas.repec.org/p/pra/mprapa/22155.html.

A.R. Gallant and D.W. Nychka. Semi-nonparametric Maximum Likelihood Estimation. *Econometrica*, 55(2):363–90, 1987.

P. Hall and Q. Yao. Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, pages 285–317, 2003.

S.W. Lee and B.E. Hansen. Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. *Econometric theory*, 10:29–29, 1994.

H. Leeb and B.M. Pötscher. Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics*, 34(5):2554–2591, 2006.

H. Leeb and B.M. Pötscher. Model selection. *Handbook of Financial Time Series*, pages 889–925, 2009.

S. Ling. Adaptive estimators and tests of stationary and nonstationary short-and long-memory ARFIMA–GARCH models. *Journal of the American Statistical Association*, 98(464):955–967, 2003.

S. Ling. Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics*, 140(2):849–873, 2007.

S. Ling and M. McAleer. On adaptive estimation in nonstationary ARMA models with GARCH errors. *The Annals of Statistics*, 31(2):642–674, 2003.

R.L. Lumsdaine. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1, 1) and covariance stationary GARCH (1, 1) models. *Econometrica*, 64(3): 575–596, 1996.

K. Mukherjee. Pseudo-likelihood estimation in ARCH models. *Canadian Journal of Statistics*, 34 (2):341–356, 2006.

A.A. Weiss. Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory*, 2(01):107–131, 1986.

2.8 Appendix

2.8.1 Technical results

Proof of Lemma 2.7.2: The first inequality in (2.7.1) is a direct extension of (4.28) and (4.29) in Francq and Zakoïan (2004). By (4.25) of the previous reference, for all $\delta > 0$ and $s \in (0,1)$, their exists a neighborhood $\mathcal{V}(\theta_0^{(r)})$ of $(\theta_0^{(r)})$ such that

$$\sup_{\theta \in \mathcal{V}(\theta_0^{(r)})} \frac{\sigma_t^2(\theta_0^{(r)})}{\sigma_t^2(\theta)} \leq K + K \sum_{i=1}^{q} \sum_{k=0}^{\infty} (1+\delta)^k \rho^{ks} \epsilon_{t-k-i}^{2s}.$$

Without loss of generality, assume that $d \geq 2$. Then

$$\left\| \sup_{\theta \in \mathcal{V}(\theta_0^{(r)})} \frac{\sigma_t^2(\theta_0^{(r)})}{\sigma_t^2(\theta)} \right\|_{\frac{d}{2}} \leq K + K \sum_{i=1}^{q} \sum_{k=0}^{\infty} (1+\delta)^k \rho^{ks} \left\| \epsilon_t^{2s} \right\|_{\frac{d}{2}}.$$

If s is chosen such as $E\left|\epsilon_t\right|^{ds} < \infty$ and if δ is chosen sufficiently small, we obtain the second inequality of (2.7.1).

We have

$$\frac{\partial l_t(\theta)}{\partial \theta} = \frac{1}{\sigma_t^2(\theta)} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \left(1 - \frac{|\epsilon_t|^r}{\sigma_t^r(\theta)} \right), \tag{2.8.1}$$

$$\begin{aligned} \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} &= \frac{1}{\sigma_t^4(\theta)} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} \left\{ \left(\frac{r}{2} + 1 \right) \frac{|\epsilon_t|^r}{\sigma_t^r(\theta)} - 1 \right\} \\ &\quad + \frac{1}{\sigma_t^2(\theta)} \frac{\partial^2 \sigma_t^2(\theta)}{\partial \theta \partial \theta'} \left(1 - \frac{|\epsilon_t|^r}{\sigma_t^r(\theta)} \right). \end{aligned} \tag{2.8.2}$$

Using the independence between η_t^* and $\sigma_t(\theta)$, the first inequality in (2.7.1) and $E\frac{|\epsilon_t|^r}{\sigma_t^r(\theta_0^{(r)})} = E|\eta_t^*|^r = 1$, we then obtain (2.7.2).

Proof of Lemma 2.7.3. The first inequality in (2.7.3) is shown by (4.6) in Francq and Zakoïan (2004), and the second one is a direct extension of (4.33) of the same paper.

The first inequality in (2.7.3) and $\inf_{\theta \in \Theta^{(r)}} \tilde{\sigma}_t^2(\theta) > 0$ imply $\sigma_t^2(\theta)/\tilde{\sigma}_t^2(\theta) \leq 1 + K\rho^t$. Thus

$$\left(\frac{\sigma_t^2(\theta)}{\tilde{\sigma}_t^2(\theta)} \right)^{d/2} \leq \left(1 + K\rho^t\right)^{d/2} \leq 1 + K\rho^t,$$

uniformly in θ, from which (2.7.4) follows.

In view of (2.8.1), letting $\tilde{\phi}_t^{(r)} = 1/\tilde{\sigma}_t^2(\theta_0^{(r)})(\partial\tilde{\sigma}_t^2(\theta_0^{(r)})/\partial\theta)$,

$$\begin{aligned}
\left\|\frac{\partial \tilde{l}_t(\theta_0^{(r)})}{\partial\theta} - \frac{\partial l_t(\theta_0^{(r)})}{\partial\theta}\right\| &= \left\|\tilde{\phi}_t^{(r)}\left(1-\frac{|\epsilon_t|^r}{\tilde{\sigma}_t^r(\theta_0^{(r)})}\right) - \phi_t^{(r)}\left(1-\frac{|\epsilon_t|^r}{\sigma_t^r(\theta_0^{(r)})}\right)\right\| \\
&= \left\|\left(1-\frac{|\epsilon_t|^r}{\tilde{\sigma}_t^r(\theta_0^{(r)})}\right)\frac{1}{\tilde{\sigma}_t^2(\theta_0^{(r)})}\left(\frac{\partial\tilde{\sigma}_t^2(\theta)}{\partial\theta} - \frac{\partial\sigma_t^2(\theta_0^{(r)})}{\partial\theta}\right)\right. \\
&\quad +\phi_t^{(r)}\left(\frac{|\epsilon_t|^r}{\sigma_t^r(\theta_0^{(r)})} - \frac{|\epsilon_t|^r}{\tilde{\sigma}_t^r(\theta_0^{(r)})}\right) \\
&\quad \left. +\phi_t^{(r)}\left(1-\frac{|\epsilon_t|^r}{\tilde{\sigma}_t^r(\theta_0^{(r)})}\right)\left(\frac{\sigma_t^2(\theta_0^{(r)})}{\tilde{\sigma}_t^2(\theta_0^{(r)})}-1\right)\right\|.
\end{aligned}$$

We then obtain

$$E\left\|\frac{\partial \tilde{l}_t(\theta_0^{(r)})}{\partial\theta} - \frac{\partial l_t(\theta_0^{(r)})}{\partial\theta}\right\| \le KE\left\{\left(1+|\eta_t^{(r)}|^r\right)\left(\frac{1}{\tilde{\sigma}_t^2(\theta_0^{(r)})}+\phi_t^{(r)}\right)\rho^t + \phi_t^{(r)}|\eta_t^{(r)}|^r\rho^t\right\} \le K\rho^t,$$

using (2.7.3)-(2.7.4), the independence between $\eta_t^{(r)}$ and $\sigma_t(\theta)$, and the existence of $E\left|\phi_t^{(r)}\right|$. We deduce (2.7.5) from the Markov inequality.

We now prove the last convergence of the lemma. From (2.7.3), we have

$$\sup_{\theta\in\mathcal{V}(\theta_0^{(r)})}\left\|\frac{1}{n}\sum_{t=1}^n\left\{\frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^{*\prime}}\right\}\right\| \le K\frac{1}{n}\sum_{t=1}^n\rho^t\Upsilon_t,$$

with $\Upsilon_t = \sup_{\theta\in\mathcal{V}(\theta_0^{(r)})}\left(\frac{|\epsilon_t|^r}{\sigma_t^r}+1\right)\left|1+\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}+\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}+\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}+\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}\right|$.
Then with the Markov inequality, the Borel-Cantelli Lemma, the Cauchy-Schwartz inequality, Assumption **A5** and (2.7.1), we can conclude.
Proof of Lemma 2.7.4. Differentiating (2.8.2), we obtain

$$\begin{aligned}
\frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} &= \frac{1}{\sigma_t^6(\theta)}\left\{2-\left(\frac{r}{2}+1\right)\left(\frac{r}{2}+2\right)\frac{|\epsilon_t|^r}{\sigma_t^r(\theta)}\right\}\frac{\partial\sigma_t^2(\theta)}{\partial\theta_i}\frac{\partial\sigma_t^2(\theta)}{\partial\theta_j}\frac{\partial\sigma_t^2(\theta)}{\partial\theta_k} \\
&+\frac{1}{\sigma_t^4(\theta)}\left\{\left(\frac{r}{2}+1\right)\frac{|\epsilon_t|^r}{\sigma_t^r(\theta)}-1\right\}\left(\frac{\partial\sigma_t^2(\theta)}{\partial\theta_i}\frac{\partial^2\sigma_t^2(\theta)}{\partial\theta_j\partial\theta_k}+\frac{\partial\sigma_t^2(\theta)}{\partial\theta_j}\frac{\partial^2\sigma_t^2(\theta)}{\partial\theta_i\partial\theta_k}+\frac{\partial\sigma_t^2(\theta)}{\partial\theta_k}\frac{\partial^2\sigma_t^2(\theta)}{\partial\theta_i\partial\theta_j}\right) \\
&+\frac{1}{\sigma_t^2(\theta)}\frac{\partial^3\sigma_t^2(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\left(1-\frac{|\epsilon_t|^r}{\sigma_t^r(\theta)}\right).
\end{aligned}$$

In view of the second inequality of (2.7.1), Assumption **A5** entails the existence of a neighborhood

$\mathcal{V}(\theta_0^{(r)})$ of $\theta_0^{(r)}$ satisfying

$$E \sup_{\theta \in \mathcal{V}(\theta_0^{(r)})} \frac{|\epsilon_t|^{2r}}{\sigma_t^{2r}(\theta)} < \infty.$$

We then conclude by Hölder's inequality and the first inequality of (2.7.1).

Lemma 2.8.1. *Under the assumptions of Lemma 2.7.1, $J^{(r)}$ is non singular and*

$$Var\left\{\frac{\partial l_t}{\partial \theta^*}(\theta_0^{(r)})\right\} = \frac{2}{r}(\mu_{2r}^* - 1)J^{(r)}. \tag{2.8.3}$$

Proof. Using the assumption **A3** and (2.8.1), we obtain

$$E\frac{\partial l_t(\theta_0^{(r)})}{\partial \theta^*} = E\left(1 - |\eta_t^*|^r\right)\ E\left\{\frac{1}{\sigma_t^2(\theta_0^{(r)})}\frac{\partial \sigma_t^2(\theta_0^{(r)})}{\partial \theta^*}\right\} = 0.$$

Thus

$$\text{Var}\left\{\frac{\partial l_t}{\partial \theta^*}(\theta_0^{(r)})\right\} = E\left\{\frac{\partial l_t}{\partial \theta^*}\frac{\partial l_t}{\partial \theta^{*\prime}}(\theta_0^{(r)})\right\} = E(1 - |\eta_t^*|^r)^2 E\left\{\frac{1}{\sigma_t^4(\theta_0^{(r)})}\frac{\partial \sigma_t^2(\theta_0^{(r)})}{\partial \theta^*}\frac{\partial \sigma_t^2(\theta^*)}{\partial \theta^{*\prime}}\right\}.$$

The non singularity of $J^{(r)}$ was proven in Francq and Zakoian (2004). □

Proof of Lemma 2.7.1. It will be convenient to introduce the notation

$$\mathbf{I}_n(\theta^*) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta^*).$$

As in the proof of Theorem 2.1 in Francq and Zakoïan (2004), we use the following steps.

a) $\lim_{n\to+\infty} \sup_{\theta^* \in \Theta^*} \left|\tilde{\mathbf{I}}_n(\theta^*) - \mathbf{I}_n(\theta^*)\right| = 0 \quad a.s.$

b) $\left(\exists t \in \mathbb{Z},\ \sigma_t^2(\theta^*) = \sigma_t^2(\theta_0^{(r)})\ a.s.\right) \Rightarrow \theta^* = \theta_0^{(r)}$.

c) $E\left|l_t(\theta_0^{(r)})\right| < \infty$ and if $\theta^* \neq \theta_0^{(r)}$, $El_t(\theta^*) > El_t(\theta_0^{(r)})$.

d) For all $\theta^* \neq \theta_0^{(r)}$, there exists a neighborhood $\mathcal{V}(\theta^*)$ such that $\liminf_{n\to\infty} \inf_{\theta \in \mathcal{V}(\theta^*)} \tilde{\mathbf{I}}_n(\theta) > El_1(\theta_0^{(r)})$

To prove a), note that by (2.7.4), almost surely

$$\begin{aligned}
\sup_{\theta^* \in \Theta^*}\left|\tilde{\mathbf{I}}_n(\theta^*) - \mathbf{I}_n(\theta^*)\right| &\leq Kn^{-1}\sum_{t=1}^{n}\sup_{\theta^* \in \Theta^*}\left\{\left|\log\left(\frac{\tilde{\sigma}_t^2(\theta^*)}{\sigma_t^2(\theta^*)}\right)\right| + \frac{|\epsilon_t|^r}{\sigma_t^r(\theta^*)}\left|\frac{\sigma_t^r(\theta^*)}{\tilde{\sigma}_t^r(\theta^*)} - 1\right|\right\} \\
&\leq Kn^{-1}\sum_{t=1}^{n}\rho^t + Kn^{-1}\sum_{t=1}^{n}|\epsilon_t|^r\rho^t.
\end{aligned}$$

We conclude with the Cesàro Lemma, using the fact that $E\left|\epsilon_t\right|^s < \infty$ for some small enough s, which is a consequence of **A2***.

Result b) was proven in Francq and Zakoian (2004). We now turn to c). We easily show that $E\left|l_t(\theta_0^{(r)})\right| < \infty$. Moreover

$$\begin{aligned} El_t(\theta^*) - El_t(\theta_0^{(r)}) &= E\left\{\log\frac{\sigma_t^2(\theta^*)}{\sigma_t^2(\theta_0^{(r)})} + \frac{2}{r}|\epsilon_t|^r\left(\frac{1}{\sigma_t^r(\theta^*)} - \frac{1}{\sigma_t^r(\theta_0^{(r)})}\right)\right\} \\ &= E\left\{\frac{2}{r}\log\frac{\sigma_t^r(\theta^*)}{\sigma_t^r(\theta_0^{(r)})} + \frac{2}{r}|\eta_t^*|^r\left(\frac{\sigma_t^r(\theta_0^{(r)})}{\sigma_t^r(\theta^*)} - 1\right)\right\} \\ &\geq 0, \end{aligned}$$

using $E|\eta_t^*|^r = 1$, the independency between η_t^* and $\sigma_t(\theta^*)$ and the inequality $\log x \leq x - 1$, the proof of c) follows. The proof of result d) is again exactly the same as in Francq and Zakoïan (2004), which completes the proof of the consistency.

Now we turn to the asymptotic normality. Using a Taylor expansion, we have

$$\begin{aligned} 0 &= n^{-1/2}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t}{\partial\theta^*}(\theta_n^*) \\ &= n^{-1/2}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t}{\partial\theta^*}(\theta_0^{(r)}) + \left\{\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t}{\partial\theta^*\partial\theta^{*\prime}}(\overline{\theta})\right\}\sqrt{n}\left(\theta_n^* - \theta_0^{(r)}\right) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\theta^*}(\theta_0^{(r)}) + J^{(r)}\sqrt{n}\left(\theta_n^* - \theta_0^{(r)}\right) + P_n + Q_n + R_n + S_n \end{aligned} \tag{2.8.4}$$

where $\overline{\theta}$ is between θ_n^* and $\theta_0^{(r)}$,

$$\begin{aligned} P_n &= \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial^2\tilde{l}_t}{\partial\theta^*\partial\theta^{*\prime}}(\overline{\theta}) - \frac{\partial^2 l_t}{\partial\theta^*\partial\theta^{*\prime}}(\overline{\theta})\right\}\sqrt{n}(\theta_n^* - \theta_0^{(r)}) \\ Q_n &= \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial^2 l_t}{\partial\theta^*\partial\theta^{*\prime}}(\overline{\theta}) - \frac{\partial^2 l_t}{\partial\theta^*\partial\theta^{*\prime}}(\theta_0)\right\}\sqrt{n}(\theta_n^* - \theta_0^{(r)}) \\ R_n &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{\partial\tilde{l}_t}{\partial\theta^*}(\theta_0^{(r)}) - \frac{\partial l_t}{\partial\theta^*}(\theta_0^{(r)})\right\} \\ S_n &= \left(J_n - J^{(r)}\right)\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^{(r)}\right) \quad \text{and} \quad J_n = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\theta^*\partial\theta^{*\prime}}(\theta_0^{(r)}). \end{aligned}$$

Using Lemma 2.7.3 and the consistency of $\widehat{\theta}_n^{(r)}$,

$$P_n = o_P\{\sqrt{n}(\theta_n^* - \theta_0^{(r)})\}, \qquad R_n = o_P(1).$$

Another Taylor expansion and Lemma 2.7.4 prove that $Q_n = o_P\{\sqrt{n}(\theta_n^* - \theta_0^{(r)})\}$. Moreover, the ergodic theorem and Lemma 2.7.2 show that $S_n = o(\sqrt{n}(\hat{\theta}_n^* - \theta_0^{(r)}))$.

Now, by the Lindeberg Central Limit Theorem (CLT) for stationary squared integrable martingales differences, and Lemma 2.8.1 we obtain

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\theta^*}(\theta_0^{(r)}) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \frac{2}{r}\left(\mu_{2r}^* - 1\right) J^{(r)}\right\}.$$

The proof of Lemma 2.7.1 is now complete.

Computation of Σ_r. First note that

$$\frac{2}{r}\left(\mu_{2r}^{(r)} - 1\right)\{J^{(r)}\}^{-1} = g(r)\{B^{(r)}\}^{-1}J^{-1}\{B^{(r)}\}^{-1}.$$

We then use Lemma 2.7.5 and the delta method to obtain (2.2.4) with

$$\Sigma_r = \Omega\begin{pmatrix} g(r)\{B^{(r)}\}^{-1}J^{-1}\{B^{(r)}\}^{-1} & \Gamma_r' \\ \Gamma_r & \Lambda_r \end{pmatrix}\Omega', \qquad \Omega = \begin{pmatrix} B & \overline{\theta}_0^{(r)} \end{pmatrix}.$$

Then,

$$\begin{aligned}
\Sigma_r &= g(r)J^{-1} + \Lambda_r\overline{\theta}_0^{(r)}\overline{\theta}_0^{(r)\prime} + \overline{\theta}_0^{(r)}\Gamma_r B + B\Gamma_r\overline{\theta}_0^{(r)\prime} \\
&= g(r)J^{-1} + \left\{2\frac{2}{r}\frac{1}{\mu_2^{(r)}}b_r + \frac{1}{\mu_2^{(r)2}}\Lambda_r\right\}\overline{\theta}_0\overline{\theta}_0' \\
&= g(r)J^{-1} + \frac{1}{\mu_2^{(r)2}}\left(2\frac{2}{r}\mu_2^{(r)}b_r + \mu_4^{(r)} - \mu_2^{(r)2} - \frac{2}{r}\mu_2^{(r)}\left(b_r + \mu_{2+r}^{(r)} - \mu_2^{(r)}\right)\right)\overline{\theta}_0\overline{\theta}_0' \\
&= g(r)J^{-1} + \frac{1}{\mu_2^{(r)2}}\left\{\frac{2}{r}\mu_2^{(r)}\left(\mu_{2+r}^{(r)} - \mu_2^{(r)} - \frac{2}{r}\mu_2^{(r)}(\mu_{2r}^{(r)} - 1)\right)\right. \\
&\qquad \left. + \mu_4^{(r)} - \mu_2^{(r)2} - \frac{2}{r}\mu_2^{(r)}\left(\mu_{2+r}^{(r)} - \mu_2^{(r)}\right)\right\}\overline{\theta}_0\overline{\theta}_0' \\
&= g(r)J^{-1} + \frac{1}{\mu_2^{(r)2}}\left\{\mu_4^{(r)} - \mu_2^{(r)2} - \left(\frac{2}{r}\right)^2\mu_2^{(r)2}(\mu_{2r}^{(r)} - 1)\right\}\overline{\theta}_0\overline{\theta}_0' \\
&= g(r)J^{-1} + \left\{\frac{\mu_4^{(r)}}{\mu_2^{(r)2}} - 1 - \left(\frac{2}{r}\right)^2(\mu_{2r}^{(r)} - 1)\right\}\overline{\theta}_0\overline{\theta}_0' \\
&= g(r)J^{-1} + \left\{\frac{\mu_4^{(r)}}{\mu_2^{(r)2}} - 1 - g(r)\right\}\overline{\theta}_0\overline{\theta}_0'.
\end{aligned}$$

The conclusion follows.

Proof of the consistency in Theorem 2.2.1. We have the almost sure convergence of $\widehat{\theta}_n^{(r)}$

to $\theta_0^{(r)}$. We need the almost sure convergence of the matrix $\widehat{B}_n^{(r)}$ to the matrix $B^{(r)}$, that is the convergence of the empirical moment $\widehat{\mu}_{2,n}^{(r)}$ to $\mu_2^{(r)}$.

With another Taylor expansion, we have

$$\frac{\epsilon_t^2}{\sigma_t^2(\widehat{\theta}_n^{(r)})} = \frac{\epsilon_t^2}{\sigma_t^2(\theta_0^{(r)})} - \frac{\epsilon_t^2}{\sigma_t^2(\widetilde{\theta})}\phi_t'(\widetilde{\theta})(\widehat{\theta}_n^{(r)} - \theta_0^{(r)}),$$

with $\widetilde{\theta}$ between $\theta_0^{(r)}$ and $\widehat{\theta}_n^{(r)}$.

With Lemma 2.7.2 and Lemma 2.7.1 and the ergodic theorem we obtain

$$(\widehat{\theta}_n^{(r)} - \theta_0^{(r)})' \frac{1}{n}\sum_{t=1}^{n} \frac{\epsilon_t^2}{\sigma_t^2(\widetilde{\theta})}\phi_t(\widetilde{\theta}) \xrightarrow{a.s.} 0.$$

Then using the strong law of large numbers, we have

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\epsilon_t^2}{\sigma_t^2(\theta_0^{(r)})} = \frac{1}{n}\sum_{t=1}^{n} \eta_t^{(r)2} \xrightarrow{a.s.} \mu_2^{(r)}.$$

2.8.2 Numerical experiments

We illustrate the convergence in distribution of the estimator of r_0, obtained in Theorem 2.3.1. In view of Remark 2.3.1, the estimation of θ_0 has no effect on the asymptotic distribution of r_n. To verify this property, and the asymptotic distribution of r_n, we simulated $N = 1,000$ sample paths of size $n = 100, n = 500$ and $n = 1,000$ of (i) $\eta_t \sim \mathcal{N}(0,1)$; (ii) $\hat{\eta}_t$ obtained from the estimation of Model (2.4.1). The smoothed densities are displayed in Figure 2.5.

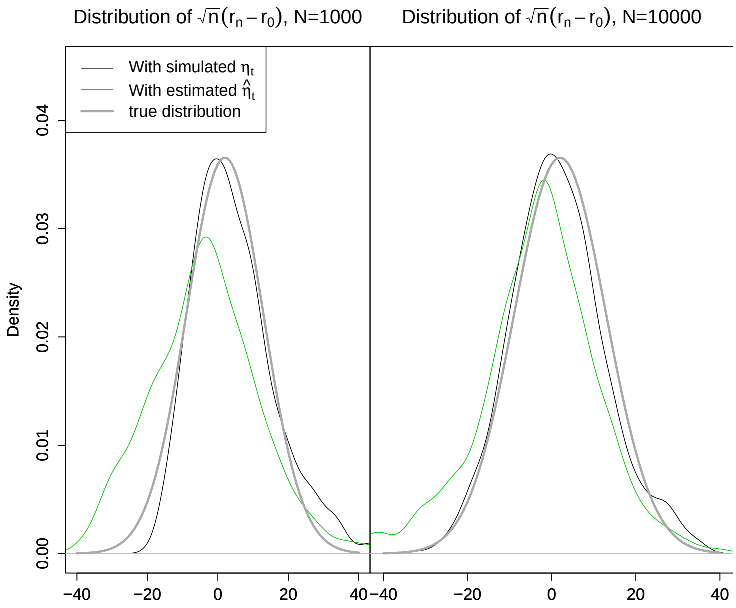


Figure 2.5: Smoothed densities for Gaussian innovations

Chapter 3

Estimating Conditionally Heteroscedastic Models with Innovations in the Domain of Attraction of a Stable Law

3.1 Introduction

ARCH models, introduced by Engle (1982) and generalized by Bollerslev (1986) are some of the most popular models for explaining financial time series. In these models, the time series is stationary but possesses a time varying conditional variance, this property can be used to explain some of the stylized facts that can be found in financial series. The GARCH modeling explains the volatility clustering but it also explains a fraction of the leptokurticity that can be found in financial time series. Empirical evidences can be found in the survey article by Shephard (1996). The most widely used estimator for the parameters of the GARCH model is the Gaussian Quasi Maximum Likelihood Estimator (QMLE). To implement this estimator, the Gaussian density is used to compute the likelihood of the model, even if the exact distribution of the error process remains unspecified. Under appropriate assumptions, the Gaussian QMLE is Consistent and Asymptotically Normal (CAN), see Berkes et al. (2003) or Francq and Zakoïan (2004).

Most of the assumptions required for the Gaussian QMLE to be CAN are mild, since one does not need to specify the true distribution of the error process, the model is less risky to be misspecified as in the Maximum Likelihood Estimation (MLE) case. The only assumption that can be challenged is the requirement that the error process possesses a finite fourth moment. The GARCH model and its derivatives are mostly applied to financial data which are known to be heavy tailed. Mandelbrot (1963) and Fama (1965) found that the unconditional distributions of most financial returns are heavy tailed and therefore do not necessarily possess a finite fourth moment. Now even if the GARCH modeling explains a part of the leptokurticity of the financial time series, the residuals are often found to remain heavy tailed. For this reason, there were several attempts to use GARCH models with non-Gaussian innovation, see Berkes and Horváth (2004) for a general approach. GARCH models with heavy tailed distributions have been studied, Bollerslev (1987) use the student t distribution and Liu and Brorsen (1995) used an α-stable distribution for the error process and studied the model empirically, see also Mittnik and Paolella (2003), Embrechts et al. (1997).

In this paper we study a stable Maximum Likelihood Estimator (MLE) of a general conditionally heteroskedastic model in which the errors follow a stable distribution. To the best knowledge of the author, the CAN property of the MLE of such a model with stable innovation has not been proven, even in the GARCH case where the model was only studied empirically. Here we prove such a result under a few assumptions about the functional form of the volatility process. By specifying the distribution of the error process (η_t) to be α-stable, we obtain a less general method than the Gaussian QMLE but we do not need any moment assumption and the model takes into account the fact the data can be heavy tailed.

The Gaussian QMLE possesses the robustness property that even if the errors are not Gaussian, provided that their distribution is in the domain of attraction of the Gaussian law, the QML estimator is still CAN. We want to obtain a similar property for the stable GARCH model. Since the only probability distributions to possess a domain of attraction are the Gaussian distribution and the family of stable laws, we use this fact to obtain a robustness property for the stable estimation. In other words, we study the asymptotic behavior of the MLE written for stable innovations when the error process is not stable but close to a stable distribution. With the Generalized Central Limit Theorem (GCLT) (see Gnedenko et al. (1968)), we can characterize the domain of attraction of a stable law. A sum of i.i.d random variables with certain properties will converge in distribution to a stable variable. If the innovation process can be written as a sum of variables, then if the sum converges, it converges in distribution toward a stable law. We use this property to give a more general result than the stable MLE. We prove that if the innovation process is not stable but converges in distribution to a stable variable, then the stable MLE (which in this case is a pseudo MLE) is still CAN.

We will study a general class a conditionally heteroscedastic model, defined by

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t = g\left(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \theta_0\right), \end{cases} \tag{3.1.1}$$

where (ϵ_t) is the observed process ($\epsilon_t \in \mathbb{R}$), (η_t) is a sequence of independent and identically (i.i.d) random variables (the error process), θ_0 is a parameter belonging to a parameter space Θ and $g: \mathbb{R}^\infty \times \Theta \mapsto \mathbb{R}_+^*$. This model contains most of the numerous derivatives of the GARCH model that have been introduced such as EGARCH, TGARCH and many others, see Bollerslev (2008) for a exhaustive (at the time) list. Model (3.1.1) contains the classical GARCH(p,q) model given by

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} b_j \sigma_{t-j}^2. \tag{3.1.2}$$

The plan of this paper is as follows. We recall useful results concerning the stable distribution in Section 2. In the third section, we study a conditional heteroscedastic model with stable innovation and prove that the MLE is stable. In section 4, we consider the case where the stable density is used to compute a pseudo MLE when the error process is not stable but converges in distribution toward a stable process. Then, we present in section 5 some simulation results and some financial applications.

3.2 Properties of stable distributions

Since the pioneer work of Mandelbrot, the class of stable distributions is commonly used in finance and in other areas such as engineering, signal processing and many other areas. There are empirical evidences that some financial processes, denoted (X_t), have regularly varying (heavy) tails, that is, $\mathbb{P}\left[X_t > x\right] \sim Kx^{-\alpha}$, when $x \to +\infty$, where K is a constant and $\alpha \in (0, 2)$. Such a process has infinite variance, therefore the standard Central Limit Theorem (CLT) cannot be applied. Fortunately, the CLT can be generalized. An iid random process (X_t) with regularly varying tails with index $\alpha < 2$ is in the domain of attraction of a stable law, that is that there exist sequences (a_n) and (b_n) such that

$$\frac{X_1 + \cdots + X_n}{a_n} - b_n \xrightarrow{\mathcal{L}} Y, \text{ where } Y \text{ is a stable law with tail parameter } \alpha.$$

Only stable distributions possess a domain of attraction. Obviously, the Gaussian law is a stable distribution since the CLT states that every random variable with finite variance is in the domain of attraction of the Gaussian law. The Gaussian distribution is a particular case of a stable distribution with $\alpha = 2$.

The formal definition of a stable variable is quite simple: non degenerate iid random variables (Z_t) are stable if there exist $a_n > 0$ and b_n such that $\frac{Z_1+\cdots+Z_n}{a_n} - b_n \stackrel{\mathcal{L}}{=} Z_1$. For a stable law, there exists, in general, no closed form of the probability density function. A stable variable is characterized by four parameters, the previously mentioned tail exponent α, a parameter of asymmetry $|\beta| \leq 1$, a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\gamma > 0$. When $\beta = 0$, the distribution is symmetric about μ. There are several special cases apart from the Gaussian case with $\alpha = 2$, where the density is explicit. A stable distribution with $\alpha = 1$ and $\beta = 0$ is a Cauchy distribution. When $\alpha = 1/2$ and $\beta = 1$, we obtain a Lévy distribution.

Though the density of a stable variable cannot be written in closed form in the general case, we can write its characteristic function, written in the (M) parametrization of Zolotarev (1986). The random variable X is called stable with parameter $\psi = (\alpha, \beta, \mu, \gamma)$ (we write $X \sim S(\psi) = S(\alpha, \beta, \mu, \gamma)$) if

$$\varphi_\psi(t) = E\left[\exp itX\right] = \begin{cases} \exp\left\{-|\gamma t|^\alpha + \gamma^\alpha i\beta \tan\left(\frac{\alpha\pi}{2}\right) t\left(|t|^{\alpha-1} - 1\right)\right\} + i\mu t, & \text{if} \quad \alpha \neq 1 \\ \exp\left\{-|\gamma t| - \gamma i\beta t\frac{2}{\pi}\log|\gamma t|\right\} + i\mu t & \text{if} \quad \alpha = 1. \end{cases} \tag{3.2.1}$$

For other parameterizations or more properties on stable distributions, see Zolotarev (1986) or Samorodnitsky and Taqqu (1994). The parametrization in (3.2.1) possesses the advantage of being continuous and differentiable with respect to all parameters, even for $\alpha = 1$. Using the inverse Fourier transform, we can express the density $f(., \psi)$ with the characteristic function

$$f(x, \alpha, \beta, \mu, \gamma) = f(x, \psi) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_\psi(t)dt. \tag{3.2.2}$$

From Bergström (1952), we give a series expansion of the stable density which will be useful to easily obtain properties of the stable distribution and to numerically compute the stable density.

Proposition 3.2.1. *For $\alpha < 1$, we have*

$$\begin{aligned} f(x, \alpha, \beta, 0, 1) &= \frac{1}{\pi}\sum_{k\geq 1}(-1)^{k+1}\frac{\Gamma(k\alpha+1)}{k!}\frac{\left(1+\tau^2\right)^{k/2}}{|x+\tau|^{k\alpha+1}} \\ &\quad \times \sin\left[k\left(\arctan\tau + \frac{\alpha\pi}{2}\right) + (k\alpha+1)\pi\mathbb{1}_{x<-\tau}\right], \end{aligned} \tag{3.2.3}$$

with $\tau = \beta\tan\frac{\alpha\pi}{2}$. For $\alpha > 1$, the latter series does not converge but the partial sum of (3.2.3) *provides an asymptotic expansion when $|x|$ tends to infinity. In this case, we have another convergent series expansion given by*

$$f(x, \alpha, \beta, 0, 1) = \frac{1}{\alpha\pi}\sum_{k\geq 0}\frac{\Gamma\left(\frac{k+1}{\alpha}\right)}{k!}x^k\cos\left[\frac{k+1}{\alpha}\arctan\tau - \frac{k\pi}{2}\right]. \tag{3.2.4}$$

This proposition can be proven as in Bergström, the parametrization differs but the idea is the same. These series expansions will be used to numerically compute the stable density. Depending of the parameters α and β and of the value of x, the series (3.2.3) or (3.2.4) will efficiently approximate the density f. If these series do not provide a good estimation, we can use the Fast-Fourier Transform (FFT) or the Laguerre quadrature, see Nolan (1997) or Matsui and Takemura (2006).

In the last proposition, the parameters μ and γ were fixed to 0 and 1. To obtain a more general formula we can use the following relation

$$f(x, \alpha, \beta, \mu, \gamma) = \frac{1}{\gamma} f\left(\frac{x-\mu}{\gamma}, \alpha, \beta, 0, 1\right). \tag{3.2.5}$$

From (3.2.2), (3.2.3) and (3.2.5), it follows that the density of a stable random variable is infinitely differentiable with respect to x, α, β and μ. From the asymptotic expansion (3.2.3), we have the tail behavior of the density f and all its derivatives. When $x \to \pm\infty$,

$$f(x, \psi) \sim K|x|^{-\alpha-1}, \tag{3.2.6}$$

$$f'(x, \psi) = \frac{\partial f}{\partial x}(x, \psi) \sim K|x|^{-\alpha-2} \tag{3.2.7}$$

$$\frac{\partial f}{\partial \alpha}(x, \psi) \sim K \log(|x|)|x|^{-\alpha-1}, \tag{3.2.8}$$

$$\frac{\partial f}{\partial \beta}(x, \psi) \sim K|x|^{-\alpha-1}, \tag{3.2.9}$$

$$\frac{\partial f}{\partial \mu}(x, \psi) \sim K|x|^{-\alpha-2}, \tag{3.2.10}$$

where K is a generic constant, which is not necessarily the same depending on whether $x \to +\infty$ or $x \to -\infty$.

The idea of using stable laws comes from the fact that only a stable variable possesses a domain of attraction. The Gaussian distribution is a particular case of stable distribution (with $\alpha = 2$), its domain of attraction contains all distributions with finite variance. The following is a CLT for heavy tailed regularly varying distribution in the particular case of a variable in the domain of normal attraction of a stable law.

Theorem 3.2.1 (Gnedenko et al. (1968), Theorem 5, § 35). *If the process (X_t) is iid with*

$$\mathbb{P}\left[X_t > x\right] \sim K_1 x^{-\alpha} \text{ when } x \to +\infty, \tag{3.2.11}$$

$$\mathbb{P}\left[X_t < x\right] \sim K_2 |x|^{-\alpha} \text{ when } x \to -\infty, \tag{3.2.12}$$

with $\alpha \in (0,2)$, $K_1 > 0$ and $K_2 > 0$, then, with

$$\beta = \frac{K_1 - K_2}{K_1 + K_2},\; a = \left\{-\alpha M(\alpha)(K_1 + K_2)\cos\frac{\alpha\pi}{2}\right\}^{1/\alpha},$$
$$M(\alpha) = \begin{cases} -\dfrac{\Gamma(1-\alpha)}{\alpha}, & \textit{when } \alpha < 1 \\ \dfrac{\Gamma(2-\alpha)}{\alpha(\alpha-1)}, & \textit{when } \alpha > 1, \end{cases}$$

we have

$$\frac{1}{an^{1/\alpha}} \sum_{t=1}^{n} (X_t - m) \xrightarrow{\mathcal{L}} Z, \tag{3.2.13}$$

where $m = EX_1$ if $\alpha > 1$, $m = 0$ if $\alpha < 1$ and $Z \sim S(\alpha, \beta, \beta \tan\frac{\alpha\pi}{2}, 1)$.

The following theorem, due to Gnedenko et al. (1968) (Theorem 2, § 46), gives a uniform version of the previous result.

Theorem 3.2.2. *Under the assumptions and with the notations of Theorem 3.2.1, if X_t has a density and if this density is of bounded variation, we have*

$$\sup_{x\in\mathbb{R}} |f_n(x) - f(x,\psi)| \to 0, \quad \textit{when } n \to +\infty, \tag{3.2.14}$$

where f_n is the density of $\frac{1}{an^{1/\alpha}} \sum_{t=1}^{n} (X_t - m)$ and $f(.,\psi)$ is the probability density of Z with $\psi = \left(\alpha, \beta, \beta\tan\frac{\alpha\pi}{2}, 1\right)$ as defined in Theorem 3.2.1.

The previous theorem has been extended by Basu and Maejima (1980) as follows.

Theorem 3.2.3. *Under the assumptions and with the notations of Theorem 3.2.2, if the characteristic function w of X_1 is such that*

$$\int_{\mathbb{R}} |w(u)|^r \, du < \infty,$$

for some integer $r \geq 1$, then for $0 \leq \delta \leq \alpha$, we have

$$\sup_{x\in\mathbb{R}} (1 + |x|)^{\delta} |f_n(x) - f(x,\psi)| \to 0, \quad \textit{when } n \to +\infty. \tag{3.2.15}$$

3.3 Conditionally heteroscedastic model with stable innovations

In this section, we study the properties of the ML estimator, for the general class of conditionally heteroscedastic models defined in (3.1.1) with a stable error process. The probability distribution of (η_t) is a stable law with parameter $\psi = (\alpha, \beta, \mu, 1)$. For identifiability reasons, the parameter γ has to be fixed to 1.

Since we work with a general model, we make some general assumptions which can be made more precise for explicit models. We will, in particular, consider the GARCH(p,q) model. We suppose,

A0 (ϵ_t) is a causal, strictly stationary and ergodic solution of (3.1.1).

Let $\epsilon_1, \cdots, \epsilon_n$ denote observations of the process (ϵ_t). The true parameter of the model is denoted $\tau_0 = (\theta_0', \psi_0')'$, where θ_0 is in $\mathbb{R}^m$ and parameterizes the known function g, $\psi_0 = (\alpha_0, \beta_0, \mu_0)'$ is the parameter of the stable density, the fixed parameter $\gamma_0 = 1$ being omitted. We still denote by $f(., \psi) = f(., \alpha, \beta, \mu)$ the density of a stable law and we also keep this notation for the stable characteristic function. The parameter τ_0 belongs to a parameter space $\Gamma = \Theta \times A \times B \times C$ such that $A \subset]0, 2[$, $B \subset]-1, 1[$ and $C \subset \mathbb{R}$.

We define the criterion, for $\tau = (\theta', \psi')' \in \Gamma$:

$$\tilde{I}_n(\tau) = \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\tau) \quad \text{with} \quad \tilde{l}_t(\tau) = \frac{1}{2} \log \tilde{\sigma}_t^2(\theta) - \log f\left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}, \psi\right),$$

where the $\tilde{\sigma}_t$ are recursively defined using some initial values and

$$\tilde{\sigma}_t^2(\theta) = g\left(\epsilon_{t-1}, \ldots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \ldots; \theta\right).$$

We also define $\sigma_t^2(\theta) = g\left(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \theta\right)$. Let τ_n be the MLE of model (3.1.1) defined by:

$$\tau_n = \underset{\tau \in \Gamma}{\operatorname{argmin}}\ \tilde{I}_n(\tau). \tag{3.3.1}$$

We define $\phi_{t,i}(\theta) = \frac{1}{\sigma_t^2(\theta)} \frac{\partial \sigma_t^2}{\partial \theta_i}(\theta)$, $\phi_{t,i,j}(\theta) = \frac{1}{\sigma_t^2(\theta)} \frac{\partial^2 \sigma_t^2}{\partial \theta_i \partial \theta_j}(\theta)$ and $\phi_{t,i,j,k}(\theta) = \frac{1}{\sigma_t^2(\theta)} \frac{\partial^3 \sigma_t^2}{\partial \theta_i \partial \theta_j \partial \theta_k}(\theta)$ and we state some assumptions on the function g and the parameter space Γ.

A1 There exists $\underline{\omega} > 0$ such that, almost surely, for any $\theta \in \Theta$, $\sigma_t(\theta) > \underline{\omega}$.

A2 For $t > 1$, $\sup_{\theta \in \Theta} \left|\sigma_t^2(\theta) - \tilde{\sigma}_t^2(\theta)\right| < K\rho^t$, where K is a constant and $0 < \rho < 1$.

A3 $\forall t,\ \sigma_t(\theta) = \sigma_t(\theta_0)$ implies $\theta = \theta_0$, a.s.

A4 The parameter space Γ is a compact set and $\tau_0 \in \Gamma$.

A5 There exists $s > 0$ such that $E|\epsilon_t|^s < +\infty$.

A6 For any compact subset Θ^* in the interior of Θ and for $(i,j,k) \in \{1,\dots,m\}$, we have

$$E \sup_{\theta\in\Theta^*} |\phi_{t,i}(\theta)| < +\infty,\ E \sup_{\theta\in\Theta^*} |\phi_{t,i,j}(\theta)| < +\infty,\ E \sup_{\theta\in\Theta^*} |\phi_{t,i}(\theta)\phi_{t,j}(\theta)| < +\infty$$
$$E \sup_{\theta\in\Theta^*} |\phi_{t,i}(\theta)\phi_{t,j,k}(\theta)| < +\infty,\ E \sup_{\theta\in\Theta^*} |\phi_{t,i,j,k}(\theta)| < +\infty,$$
$$E \sup_{\theta\in\Theta^*} |\phi_{t,i}(\theta)\phi_{t,j}(\theta)\phi_{t,k}(\theta)| < +\infty.$$

A7 For $t>1$, $\sup_{\theta\in\Theta}\left\|\frac{\partial\tilde\sigma_t^2}{\partial\theta}(\theta) - \frac{\partial\sigma_t^2}{\partial\theta}(\theta)\right\| < K\rho^t$, and $\sup_{\theta\in\Theta}\left\|\frac{\partial^2\tilde\sigma_t^2}{\partial\theta\partial\theta'}(\theta) - \frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}(\theta)\right\| < K\rho^t$.

A8 The components of $\frac{\partial\sigma_t^2}{\partial\theta}(\tau)$ are linearly independent.

We prove that the estimator τ_n is CAN, the first result establishes the consistency, then with additional assumptions, we obtain the asymptotic normality of the estimator.

Theorem 3.3.1. *Under Assumptions **A0-A5**, the estimator τ_n is consistent,*

$$\tau_n \underset{n\to+\infty}{\longrightarrow} \tau_0 \quad a.s. \tag{3.3.2}$$

*If, in addition **A6-A8** hold,*

$$\sqrt{n}(\tau_n - \tau_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, J^{-1}\right), \tag{3.3.3}$$

with $J = E\left[\frac{\partial^2 l_t(\tau_0)}{\partial\tau\partial\tau'}\right] = E\left[\frac{\partial l_t}{\partial\tau}(\tau_0)\frac{\partial l_t}{\partial\tau'}(\tau_0)\right]$, *where* $l_t(\tau) = \frac{1}{2}\log\sigma_t^2(\theta) - \log f\left(\frac{\epsilon_t}{\sigma_t(\theta)}, \psi\right)$.

Remark 3.3.1. The numerous assumptions of this theorema re due to the fact that Model (3.1.1) is very general. For more specific formulations, some of theses assumptions vanish. For exemple in the case of the GARCH(p,q) model of Equation (3.1.2), Assumptions **A1, A2, A3, A5, A6, A7** and **A8** are obtained in Francq and Zakoïan (2004).

Concerning Assumption **A0**, in the case of the GARCH(p,q), we require the top Lyapunov exponent associated to the model (see for instance Berkes et al. (2003)) to be strictly negative. In the case $p=q=1$, we draw the stationarity zones for the parameters a and b for different values of α. Here, we use a symmetric stable distribution (i.e. $\beta=0$). In Figure 3.1, we numerically obtained the strict stationarity zones which for each α, is the area under the curve. If $\alpha=2$, this is the stationarity zone for a GARCH(1,1) model with Gaussian innovation but in the case $\alpha<2$, the strict stationarity zone becomes smaller as α decreases. This can be explained by the fact that the smaller α, the thicker the tails. Then if the parameters a and b take too large values, the persistence of σ_t is too strong and σ_t explodes to infinity.

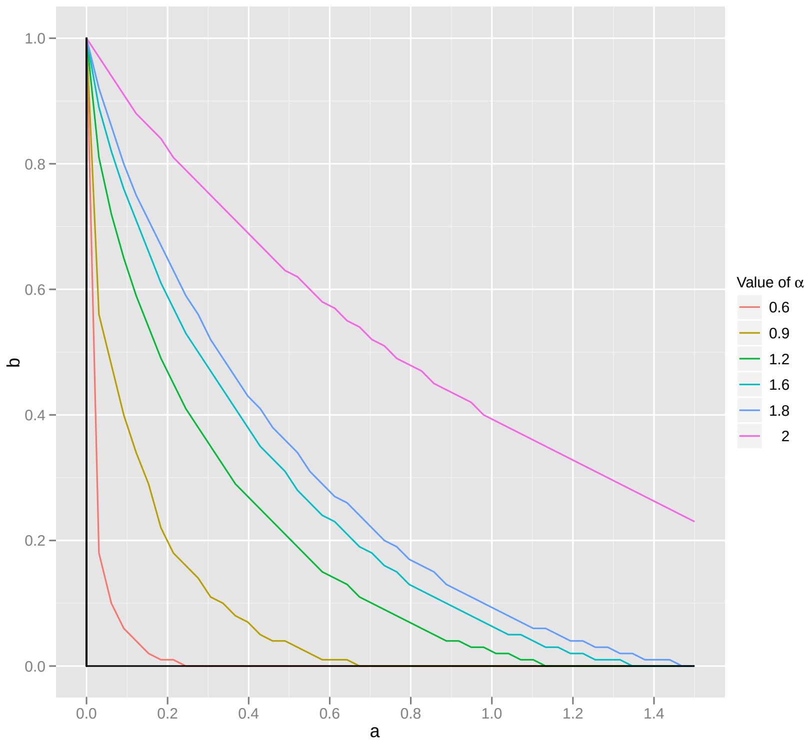


Figure 3.1: Strict stationarity zones for a GARCH(1,1) model with α-stable innovation. The curves correspond, in decreasing order to $\alpha = 2$, ..., $\alpha = 0.6$.

3.4 When the innovation process converges in distribution to a stable distribution

In this section, for clarity purpose, we will enunciate the results for a GARCH(p,q) model (Model (3.1.2)). The same results could be obtained for a more general model but at the cost of some technical assumptions on the function g. We write a different version of Model (3.1.2) with an innovation process (η_{nt}) which now depends on n. We have,

$$\begin{cases} \epsilon_{nt} = \sigma_{nt}\eta_{nt} \\ \sigma_{nt}^2 = \omega_0 + \sum_{i=1}^{q} a_{0i}\epsilon_{nt-i}^2 + \sum_{j=1}^{p} b_{0j}\sigma_{nt-j}^2, \quad \forall t \in \mathbb{Z}, \ \forall n \in \mathbb{N}, \end{cases} \tag{3.4.1}$$

where the process $(\eta_{nt})_t$ is iid with p.d.f. f_n and converges in distribution toward a stable variable with parameter $\psi_0 = (\alpha_0, \beta_0, \mu_0)'$. This assumption will be made explicit below. As in Section 3.3, the parameter γ_0 is omitted and fixed to 1 for identifiability reasons. The true parameter of the model is $\tau_0 = (\theta_0', \psi_0')'$, where $\theta_0 = (\omega_0, a_{01}, \ldots, a_{0q}, b_{01}, \ldots, b_{0p})'$ belongs to a parameter space $\Theta \subset (0, +\infty) \times [0, +\infty)^{p+q}$. The parameter τ_0 belongs to $\Gamma = \Theta \times A \times B \times C$, with A, B, C

as in Section 3.3. Let the polynomials $\mathcal{A}_\theta(z) = \sum_{i=1}^q a_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^p b_j z^j$ where $\theta = (\omega, a_1, \dots, a_q, b_1, \dots, b_p)'$. For θ such that $\sum_{j=1}^p b_j < 1$ and $b_j \geq 0$ for $j \in \{1, \dots, p\}$, define the function $\sigma_{nt}^2(\theta) = \frac{\omega}{\mathcal{B}_\theta(1)} + \mathcal{B}_\theta^{-1}(L)\mathcal{A}_\theta(L)\epsilon_{nt}^2$, where L denotes the lag operator.

We suppose that the process (η_{nt}) is iid for every $n \in \mathbb{N}$, but we need a stronger assumption. Define for $t \in \mathbb{Z}$, $\mathcal{F}_t = \sigma\left(\cup_{n\in\mathbb{N}}\mathcal{F}_{nt}\right)$, where $\mathcal{F}_{nt} = \sigma\left\{\eta_{nu}; u \leq t-1\right\}$ and suppose that for any $t \in \mathbb{Z}$ and for any $n \in \mathbb{N}$, η_{nt} is independent of $\mathcal{F}_t$.

We define a pseudo maximum likelihood estimator. The density of (η_{nt}) is not specified but we suppose the convergence of this process to the stable distribution with p.d.f. $f(., \psi_0)$, where ψ_0 is a unknown parameter. We use this density to build a pseudo MLE. We have for $\tau \in \Gamma$,

$$\tilde{I}_n(\tau) = \frac{1}{n}\sum_{t=1}^n \tilde{l}_{nt}(\tau) \quad \text{with} \quad \tilde{l}_{nt}(\tau) = \log\tilde{\sigma}_{nt}(\theta) - \log f\left(\frac{\epsilon_{nt}}{\tilde{\sigma}_{nt}(\theta)}, \psi\right),$$

where the $(\tilde{\sigma}_{nt}(\theta))_t$ are recursively defined using some initial values. We define

$$\tau_n = \underset{\tau\in\Gamma}{\operatorname{argmin}}\ \tilde{I}_n(\tau). \tag{3.4.2}$$

We have kept the same notations as in the previous section because all the involved quantities are defined in the same way and play the same role. The objects of this section simply display an additional n subscript.

We define γ_n as the top Lyapunov exponent associated to Model (3.4.1) and γ as the top Lyapunov exponent associated to the model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega_0 + \sum_{i=1}^q a_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^p b_{0j}\sigma_{t-j}^2, \quad \forall t \in \mathbb{Z}, \end{cases} \tag{3.4.3}$$

with $\eta_t \sim S(\psi_0)$. The top Lyapunov exponent will be more precisely defined in the Proofs section.

B1 $\tau_0 \in \Gamma$ and Γ is a compact.

B2 $\gamma < 0$ and $\forall \theta \in \Theta$, $\sum_{j=1}^p b_j < 1$.

B3 There exists $\delta > 1$ such that for any $n \in \mathbb{N}$, $E|\eta_{nt}|^\delta < +\infty$ and $\sup_{x\in\mathbb{R}} (1 + |x|)^\delta |f_n(x) - f(x, \psi_0)| \to 0$.

B4 If $p > 0$, $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ have no common roots, $\mathcal{A}_{\theta_0}(1) \neq 0$ and $a_{0q} + b_{0p} \neq 0$.

B5 We have $\sup_{n\in\mathbb{N}} \alpha_{\epsilon_n}(h) \leq K\rho^h$, where $\alpha_{\epsilon_n}(h)$ for $h \in \mathbb{N}$ is the strong mixing coefficient of the process (ϵ_{nt}).

B6 $\tau_0 \in \overset{\circ}{\Gamma}$, where $\overset{\circ}{\Gamma}$ denotes the interior of Γ.

Theorem 3.4.1. *Under Assumptions **B1-B5**, the estimator τ_n is consistent,*

$$\tau_n \underset{n\to+\infty}{\longrightarrow} \tau_0, \ a.s. \tag{3.4.4}$$

*If, in addition **B6** holds,*

$$\sqrt{n}(\tau_n - \tau_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, J^{-1}\right), \tag{3.4.5}$$

with $J = E\left[\frac{\partial^2 l_t(\tau_0)}{\partial\tau\partial\tau'}\right] = E\left[\frac{\partial l_t}{\partial\tau}(\tau_0)\frac{\partial l_t}{\partial\tau'}(\tau_0)\right]$*, where $l_t(\tau)$ is defined in the proofs.*

Remark 3.4.1. The variance-covariance matrix in Theorem 3.4.1 is the same as in Theorem 3.3.1. There is asymptotically no cost for not specifying the true distribution of the innovation and instead assuming that the process converges in distribution to a stable law.

Remark 3.4.2. The required assumptions for this result are very mild, **B1, B2, B4** and **B6** are also needed for the classical Gaussian QML. Assumption **B3** is specific to the problem and is verified in the case described hereafter where the innovation can be written as the sum of an iid process,

$$\eta_{nt} = \frac{1}{k_n^{1/\alpha}} \sum_{i=1}^{k_n} \nu_{it},$$

where for any $t \in \mathbb{Z}$, $(\nu_{it})_i$ is iid and where $(k_n)_n$ is an increasing sequence of integers and with $\alpha \in (1, 2)$ such that there exist K_1 and K_2 such that

$$\mathbb{P}\left[\nu_{1t} > x\right] \sim K_1 x^{-\alpha}, \text{ when } x \to +\infty$$
$$\mathbb{P}\left[\nu_{1t} < x\right] \sim K_2 x^{-\alpha}, \text{ when } x \to -\infty,$$

then if the density of ν_{it} satisfies the assumptions of Theorems 3.2.2 and 3.2.3, the Assumption **B3** is verified with Theorem 3.2.3.

Remark 3.4.3. In the Gaussian QML case, the asymptotic inverse variance-covariance matrix J depends on the unobserved distribution of the process (η_t). Here the matrix J depends on the limit in distribution of the innovation process (η_{nt}). We can define an estimator for the matrix J, based on the process (ϵ_{nt}) and prove that this estimator is consistent.

Remark 3.4.4. About Assumption **B5**, for each value of n, the fact that there exist constants

K and ρ such that $\alpha_{\epsilon_n}(h) \leq K\rho^h$ has been proved by Boussama (1998). We only assume that this is also true for $\sup_{n\in\mathbb{N}} \alpha_{\epsilon_t}(h)$.

Theorem 3.4.2. *Define* $J_n = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \tilde{l}_{nt}}{\partial\tau\partial\tau'}(\tau_n)$. *With the assumptions of Theorem 3.4.1, we have*

$$J_n \underset{n\to+\infty}{\longrightarrow} J, \quad a.s. \tag{3.4.6}$$

3.5 Numerical experiments

In this section, we describe a simulation experiment which aims at studying the behavior of the pseudo MLE for finite samples, and for an innovation process whose distribution is close to a stable distribution. We use the algorithm of Chambers et al. (1976) to simulate stable processes and Proposition 3.2.1 to compute the stable density.

We want to verify that even if the model is misspecified, that is if we use a stable MLE when the true distribution of the innovation process is not stable, the GARCH coefficients are still correctly estimated. We use a Student distribution with degree of freedom α (which by Theorem 3.2.1 is in the domain of attraction of a stable law of parameter α) to build an innovation process of the form

$$\eta_t^{(K)} = \frac{1}{K^{1/\alpha}} \sum_{k=1}^{K} \nu_{k,t}, \quad \text{with} \quad (\nu_{k,t})_k \overset{iid}{\sim} t_\alpha.$$

Using the results of Section 3.2, we obtain that, when K tends to infinity, $\eta_t^{(K)}$ converges in distribution toward an alpha stable law. The problem is that, for identifiability reason, the parameter γ of the stable distribution cannot be estimated and has to be fixed to 1. When $K = +\infty$, the process $\left(\eta_t^{(+\infty)}\right)_t$ is alpha stable with parameter $\psi = (\alpha, \beta, \mu, \gamma)'$. The parameter ψ depends on the degree of freedom of the Student process $(\nu_{k,t})_k$ and can be calculated. For a generic case, we have $\gamma \neq 1$. If we estimate a GARCH model with innovation process $\left(\eta_t^{(+\infty)}\right)_t$ using a stable pseudo MLE method, we would obtain estimates of the parameters of the same model but written under a different identifiability assumption. For example, if we aim to estimate the model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t^{(+\infty)} \\ \sigma_t^2 = \omega_0 + a_0\epsilon_{t-1}^2 + b_0\sigma_{t-1}^2, \end{cases}$$

the stable pseudo MLE defined in the previous sections will converge toward $\theta_0^* = \left(\frac{\omega_0}{\gamma^2}, \frac{a_0}{\gamma^2}, b_0\right)'$, see Francq et al. (2011) for more details on reparametrization of GARCH models. Note that the estimation of the "GARCH" parameter b_0 is not affected by the identification problem. In order to compare estimates of the same quantity, it is thus important that the model is similarly identified for each value of K. Thereafter, we use the following identifiability condition.

- If the innovation process $(\eta_t)_t$ is stable, then it is stable with parameter $\gamma_0 = 1$ (we recall that if X is stable with parameter $\gamma = \gamma_0 > 0$ then $\frac{X}{\gamma_0}$ is stable with parameter $\gamma = 1$).

– It the innovation process $(\eta_t)_t$ is not stable, we require that, among the family of stable distributions, the closest distribution to the distribution of $(\eta_t)_t$ in the sense of the Kulback-Leibler distance is stable with parameter $\gamma_0 = 1$.

Thus, for each K, we estimate the quantity j_K, defined such that the innovation process defined by $\eta_{t,K} = \frac{1}{j_K K^{1/\alpha}} \sum_{k=1}^{K} \nu_{k,t}$ satisfies the identifiability assumption. It is important to note that if we use another normalizing constant than j_K, the results of the estimation by stable pseudo ML are as efficient as in the case where we use j_K, the model is simply written under a different identifiability condition.

We generated 1000 samples of size $n = 1000$ for different values of K ($K = 500,\ K = 1000,\ K = 10000,\ K = 100000$ and $K = \infty$) of the following model and estimate its parameters by stable MLE (or pseudo MLE).

$$\begin{cases} \epsilon_t = \sigma_t \eta_{t,K} \\ \sigma_t^2 = 0.01 + 0.02\epsilon_{t-1}^2 + 0.7\sigma_{t-1}^2 \\ \eta_{t,K} = \dfrac{1}{j_K K^{1/\alpha}} \displaystyle\sum_{k=1}^{K} \nu_{k,t}. \end{cases} \tag{3.5.1}$$

We can summarize the simulation scheme with the following steps. For a parameter θ_0, for $K > 0$ and for a student distribution of degree α,

– Step 1: we simulate 1000 samples of the variable $\left(\frac{1}{K^{1/\alpha}} \sum_{k=1}^{K} \nu_{k,t}\right)_t$, then we fit a stable distribution on each sample. For each sample s, we denote by $\psi_s = (\alpha_s, \beta_s, \mu_s, \gamma_s)'$ the results of this estimation.
– Step 2: we compute $j_K = \frac{1}{1000} \sum_{s=1}^{1000} \gamma_s$.
– Step 3: we draw 1000 samples of Model (3.5.1).
– Step 4: for each sample s, we estimate $\tau_n^{(s)} = \left(\theta_n^{(s)'}, \psi_n(s)'\right)'$ by stable PMLE.

The results of these estimations are presented in Table 3.1. For each of the six parameters (three for the GARCH dynamic and three for the stable distribution), we give the quotient of the Root Mean Squared Errors (RMSE) of the corrected stable pseudo MLE and of the RMSE of the MLE, corresponding to the case $K = \infty$. This statistic is given by $Q_i^K = \frac{RMSE_i^{K=+\infty}}{RMSE_i^{K=K}}$ where the RMSE for K can be obtained by, for $i \in \{1, \cdots, 6\}$, $RMSE_i^K = \frac{1}{1000} \sum_{s=1}^{1000} \left(\theta_{n,i}^{(s)} - \theta_{0,i}\right)^2$. The greater Q_i is, the better the stable pseudo MLE is with respect to the MLE. In this simulation framework, we do not compare different methods of estimation. We use the same method but applied to different data generating processes. We can see that in most cases, the different $(Q_i)_i$ increase with K and that the RMSE of the misspecified model is in general quite close the RMSE of the asymptotic case. The behavior of our estimator is not much affected by the specification error on the density used to compute the likelihood.

	$\alpha = 1.6$				$\alpha = 1.4$			
	500	1000	10000	100000	500	1000	10000	100000
w	0.79	0.66	0.89	0.86	0.80	0.83	0.78	0.91
a	0.50	0.54	0.82	0.83	0.85	0.74	0.81	0.97
b	0.76	0.76	0.94	0.99	0.97	0.93	0.94	0.97
α	0.58	0.70	0.91	1.03	0.93	0.94	0.95	0.96
β	1.27	1.22	1.05	1.01	0.84	1.12	0.86	1.04
δ	0.96	0.87	0.89	0.86	0.90	1.09	0.90	0.93

Table 3.1: Ratio of RMSE for the parameters of the model for several values of K and α.

3.6 Application to financial data

In this section, we consider the daily returns of several indices and currency rates, namely the EURUSD, JPYUSD, DJA, DJI, DJT, DJU, CAC, FTSE, NIKKEI, DAX, S&P50 and the SMI. A GARCH(1,1) model with stable innovations is estimated on each of these series. The samples extend from January 1, 2008 to December 31, 2010. The estimated α's are lower for the period before 2008 so we only kept three years of data. Table 3.2 shows the results of these estimations with the standard deviation in parenthesis. We can see that even if the GARCH modeling explain a fraction of the leptokurticity of the series, the residuals still possess heavy tails since in most cases α is around 1.8 and thus different from 2 (except for the NIKKEI). When $\alpha = 2$, the parameter β cannot be identified.

Index	$\omega \times 10^5$		$a \times 10^2$		b		α		β		μ	
EURUSD	0.133	(0.047)	4.120	(0.720)	0.877	(0.025)	1.900	(0.020)	-0.007	(0.480)	-0.007	(0.075)
JPYUSD	0.578	(0.270)	3.370	(0.770)	0.667	(0.130)	1.780	(0.037)	-0.148	(0.190)	-0.137	(0.062)
INRUSD	0.042	(0.016)	2.300	(0.630)	0.912	(0.024)	1.820	(0.077)	0.242	(0.230)	-0.016	(0.067)
DJA	0.050	(0.044)	4.610	(0.540)	0.895	(0.014)	1.820	(0.110)	-0.039	(0.280)	0.028	(0.071)
DJI	0.046	(0.038)	4.670	(0.400)	0.893	(0.011)	1.840	(0.083)	-0.501	(0.270)	0.114	(0.071)
DJT	0.071	(0.070)	3.820	(0.560)	0.916	(0.014)	1.880	(0.087)	-0.060	(0.330)	0.060	(0.067)
DJU	0.095	(0.045)	4.690	(0.700)	0.887	(0.018)	1.910	(0.068)	-0.296	(0.360)	0.040	(0.064)
CAC40	0.221	(0.092)	2.970	(0.550)	0.914	(0.016)	1.880	(0.072)	-0.220	(0.210)	-0.010	(0.058)
FTSE	0.156	(0.067)	3.660	(0.610)	0.898	(0.019)	1.820	(0.050)	-0.222	(0.230)	0.059	(0.060)
NIKKEI	0.376	(0.140)	5.700	(0.490)	0.863	(0.021)	2.000	(0.090)	NA	(NA)	-0.021	(0.053)
DAX	0.136	(0.051)	2.830	(0.510)	0.922	(0.014)	1.860	(0.079)	-0.252	(0.270)	0.034	(0.065)
S&P500	0.909	(0.250)	3.870	(0.780)	0.766	(0.055)	1.770	(0.057)	-0.162	(0.180)	0.103	(0.062)
SMI	0.112	(0.049)	4.850	(0.470)	0.876	(0.015)	1.850	(0.077)	-0.255	(0.250)	0.028	(0.064)

Table 3.2: Estimation of a GARCH(1,1), standard errors in parenthesis.

These estimations can be used to compute Value-at-Risk (VaR). If ϵ_t is the return of the series, the VaR with coverage probability p at time t is defined as the quantity $\mathrm{VaR}_t(p)$ satisfying

$$\mathbb{P}_t[\epsilon_t \leq \mathrm{VaR}_t(p)] = p,$$

where $\mathbb{P}_t$ is the probability measure conditionally to the time $t-1$ information set. Using a conditionally heteroscedastic model with stable innovation, if $\tau_n = (\theta_n', \psi_n')'$ is the MLE (or pseudo MLE if the innovation is not stable but assumed to be close to a stable distribution), we have

$$\mathrm{VaR}_t(p) = \tilde{\sigma}_{t-1}(\theta_n) F^{\leftarrow}(p, \psi_n),$$

where $F^{\leftarrow}(., \psi)$ is the quantile function of a stable distribution of parameter ψ (with $\gamma = 1$). We compare this *stable* VaR to the VaR computed using a GARCH(1,1) model, estimated with the Gaussian QMLE. We compute a Gaussian QMLE on the indices used in Table 3.2 and obtain the gaussian counterparts of the parameters in this table.

Then we compute the Gaussian VaR and the stable VaR on an outsample data set (from January 1, 2011 to January 31, 2012). We give the results for the VaR of level 1% and 5% in Table 3.3.

Index	Level=0.01		Level=0.05	
	Stable	Gaussian	Stable	Gaussian
EURUSD	0.0108	0.0180	0.0755	0.0791
JPYUSD	0.0035	0.0035	0.0177	0.0213
INRUSD	0.0106	0.0142	0.0426	0.0355
DJA	0.0221	0.0221	0.0551	0.0515
DJI	0.0147	0.0221	0.0551	0.0588
DJT	0.0294	0.0331	0.0588	0.0478
DJU	0.0110	0.0110	0.0551	0.0588
CAC40	0.0108	0.0179	0.0645	0.0717
FTSE	0.0037	0.0110	0.0625	0.0662
NIKKEI	0.0114	0.0114	0.0342	0.0342
DAX	0.0072	0.0143	0.0717	0.0789
S&P500	0.0074	0.0221	0.0699	0.0588
SMI	0.0109	0.0181	0.0688	0.0688
Mean	**0.0118**	**0.0168**	**0.0563**	**0.0563**

Table 3.3: Frequency of hits for the Gaussian VaR and the stable VaR

The two methods give very close results for $p = 0.05$, but the Gaussian method seems unable to explain the extremes of the distribution and the stable VaR seems to give better results for $p = 0.01$. In this case and for almost every index, the Gaussian VaR is underestimated. There are too many hits in the sample. We can conclude that the residuals of the GARCH model estimated by Gaussian QMLE are too leptokurtic to be explained by a Gaussian distribution. To conclude, the stable distribution seems to do a better job to explain the tails of the studied financial series.

3.7 Proofs

Throughout the proofs and the paper, we denote by K and ρ generic constants whose values $K > 0$ and $0 < \rho < 1$ can vary from line to line.

3.7.1 Proof of the consistency in Theorem 3.3.1

Let I_n (resp. l_t) be the equivalent of $\tilde{I}_n$ (resp. $\tilde{l}_t$) when an infinite past is known,

$$I_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} l_t(\tau) \quad \text{with} \quad l_t(\tau) = \frac{1}{2}\log\sigma_t^2(\theta) - \log f\left(\frac{\epsilon_t}{\sigma_t(\theta)}, \psi\right).$$

We first prove that the initial values are asymptotically negligible, that is

$$\lim_{n\to+\infty}\sup_{\tau\in\Gamma}\left|I_n(\tau) - \tilde{I}_n(\tau)\right| = 0 \quad \text{a.s.} \tag{3.7.1}$$

We have,

$$\sup_{\tau\in\Gamma}\left|I_n(\tau) - \tilde{I}_n(\tau)\right| \le \frac{1}{n}\sum_{t=1}^{n}\sup_{\tau\in\Gamma}\left|\frac{1}{2}\log\frac{\sigma_t^2(\theta)}{\tilde{\sigma}_t^2(\theta)} + \log\frac{f(\tilde{\eta}_t(\theta),\psi)}{f(\eta_t(\theta),\psi)}\right|,$$

where $\tilde{\eta}_t(\theta) = \frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}$ and $\eta_t(\theta) = \frac{\epsilon_t}{\sigma_t(\theta)}$.

The function f is infinitely differentiable with respect to x, therefore, for $\tau = (\theta', \psi')' \in \Gamma$, we have

$$\left|\log f(\tilde{\eta}_t(\theta),\psi) - \log f(\eta_t(\theta),\psi)\right| < \sup_{x\in\mathbb{R}}\left|\frac{f'(x,\psi)}{f(x,\psi)}\right| \left|\tilde{\eta}_t(\theta) - \eta_t(\theta)\right|.$$

Next, using the asymptotic expansion (3.2.6)-(3.2.7), we obtain that $\frac{f'(x,\psi)}{f(x,\psi)}$ tends to 0, when $|x|$ tends to infinity and thus that $x \mapsto \frac{f'(x,\psi)}{f(x,\psi)}$ is bounded on $\mathbb{R}$. Under Assumption **A4**, we obtain $\sup_{\tau\in\Gamma}\sup_{x\in\mathbb{R}}\left|\frac{f'(x,\psi)}{f(x,\psi)}\right| < \infty$. We have using Assumption **A1** and **A2**, $\sup_{\tau\in\Gamma}|\tilde{\eta}_t(\theta) - \eta_t(\theta)| < K|\epsilon_t|$ and $\sup_{\tau\in\Gamma}|\tilde{\sigma}_t(\theta) - \sigma_t(\theta)| < K|\epsilon_t|\rho^t$. Thus we have $\sup_{\tau\in\Gamma}|\log f(\tilde{\eta}_t(\theta),\tau) - \log f(\eta_t(\theta),\tau)| < K|\epsilon_t|\rho^t$ and finally

$$\sup_{\tau\in\Gamma}\left|I_n(\tau) - \tilde{I}_n(\tau)\right| < \frac{1}{n}\sum_{t=1}^{n} K|\epsilon_t|\rho^t.$$

With the Markov inequality, the Borel-Cantelli Lemma, the existence of a moment of order s for the processus (ϵ_t) (Assumption **A5**), and Assumption **A0**, we obtain that $|\epsilon_t|\rho^t$ converges to 0 almost surely when t tends to infinity. Then, using the Cesàro Lemma, we obtain (3.7.1).

We now prove that $El_t(\tau) > El_t(\tau_0)$, for $\tau \neq \tau_0$.

$$\begin{aligned} El_t(\tau) - El_t(\tau_0) &= E\left[\log \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \frac{f(\eta_t(\theta),\psi)}{f(\eta_t,\psi_0)}\right] \\ &\leq E\left[\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \frac{f(\eta_t(\theta),\psi)}{f(\eta_t,\psi_0)}\right] - 1 \\ &= E\left[E\left[\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \frac{f(\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\eta_t,\psi)}{f(\eta_t,\psi_0)} |\mathcal{G}_t\right]\right] - 1 \\ &= 0. \end{aligned}$$

To obtain the last equality, we used the fact that $\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}$ is in $\mathcal{G}_t = \sigma\left\{\eta_u;\ u \leq t-1\right\}$.

Now, we show that, if $El_t(\tau) = El_t(\tau_0)$, then $\tau = \tau_0$ a.s. We have $\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \frac{f(\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\eta_t,\psi)}{f(\eta_t,\psi_0)} = 1$ a.s. Let $a_{t-1} = \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}$, since (η_t) has a Lebesgue density, we have

$$\forall x \in \mathbb{R},\ a_{t-1} f(a_{t-1}x,\psi) = f(x,\psi_0) \quad a.s. \tag{3.7.2}$$

We define X as a stable variable with parameters ψ_0, $X \sim S(\psi_0)$, and $Y = a_{t-1}X$. Then, using (3.7.2) we show that the pdf of Y conditionally to $\mathcal{G}_t$ is $f(x,\psi)$, thus $Y \sim S(\psi)$. Now, for $u \in \mathbb{R}$, we write the characteristic function of Y and obtain

$$E\left[e^{iuY}|\mathcal{G}_t\right] = E\left[e^{iua_{t-1}X}|\mathcal{G}_t\right],$$

that is $\varphi_\psi(u) = \varphi_{\psi_0}(a_{t-1}u)$. Applying the modula to the previous equation, we obtain

$$\forall u \in \mathbb{R},\ \exp\left\{-|u|^\alpha\right\} = \exp\left\{-|a_{t-1}u|^{\alpha_0}\right\}.$$

Therefore we have $\alpha = \alpha_0$ and we easily obtain $\beta = \beta_0$ and $\mu = \mu_0$. We also obtain $a_{t-1} = 1$ almost surely and we deduce with Assumption **A3** that $\theta = \theta_0$ a.s.

Now, for $\tau \in \Gamma$, let $V_k(\tau)$ be the open ball with center τ and radius $1/k$, using (3.7.1)

$$\liminf \inf_{\tau^* \in V_k(\tau)\cap\Gamma} \tilde{I}_n(\tau^*) \geq \liminf \inf_{\tau^* \in V_k(\tau)\cap\Gamma} I_n(\tau^*), \tag{3.7.3}$$

The ergodic theorem yields $\liminf \inf_{\tau^* \in V_k(\tau)\cap\Gamma} I_n(\tau^*) = E\left[\inf_{\tau^* \in V_k(\tau)\cap\Gamma} l_t(\tau^*)\right]$. Now when k increases to ∞, $E\left[\inf_{\tau^* \in V_k(\tau)\cap\Gamma} l_t(\tau^*)\right]$ increases to $E\left[l_t(\tau)\right]$. Therefore, we have

$$\forall \tau \neq \tau_0,\ \exists V(\tau),\ \liminf \inf_{\tau^* \in V_k(\tau)\cap\Gamma} \tilde{I}_n(\tau^*) > El_t(\tau_0).$$

We conclude by a standard compactness argument, using Assumption **A4** and obtain (3.3.2).

3.7.2 Proof of the asymptotic normality in Theorem 3.3.1

Lemma 3.7.1. *Under the assumptions of Theorem 3.3.1, we have*

$$\sqrt{n}\frac{\partial I_n}{\partial \tau}(\tau_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, J). \tag{3.7.4}$$

Proof. For $\lambda \in \mathbb{R}^{m+3}$, $n > 0$ and $t > 0$, let $\nu_{nt} = \frac{1}{\sqrt{n}}\lambda' \frac{\partial l_t}{\partial \tau}(\tau_0)$. We prove that $(\nu_{nt}, \mathcal{G}_{t-1})$ is a martingale difference. We have, for $\tau = (\theta', \psi')' \in \Gamma$

$$\frac{\partial l_t}{\partial \theta}(\tau) = \frac{1}{2}\phi_t(\theta) Z_t(\tau) \tag{3.7.5}$$

$$\frac{\partial l_t}{\partial \psi}(\tau) = -\frac{\partial \log f}{\partial \psi}(\eta_t(\theta), \psi), \tag{3.7.6}$$

with $Z_t(\tau) = 1 + \eta_t(\theta)\frac{f'(\eta_t(\theta),\psi)}{f(\eta_t(\theta),\psi)}$.

Since $\sigma_t^2(\theta_0) \in \mathcal{G}_t$, $\sigma_t(\theta_0)$ and η_t are independent and

$$E\left|\frac{\partial l_t}{\partial \theta}(\tau_0)\right| = E\left|\phi_t(\theta_0)\right| E\left|Z_t({}_a u_0)\right|.$$

The function $x \mapsto 1 + x\frac{f'(x,\psi_0)}{f(x,\psi_0)}$ is bounded, therefore we have $E\left|Z_t(\tau_0)\right| < +\infty$. Moreover, with Assumption **A6**, we have $E\left|\phi_t(\theta_0)\right| < +\infty$. With (3.2.8)-(3.2.10), we obtain $E\left|\nu_{nt}\right| < +\infty$.

Now, we have

$$E\left[\frac{\partial l_t}{\partial \theta}(\tau_0)|\mathcal{G}_t\right] = \phi_t(\theta_0) E\left[Z_t(\tau_0)\right].$$

And,

$$E\left[Z_t(\tau_0)\right] = 1 + \int_{\mathbb{R}} x f'(x, \psi_0) dx = 0.$$

We have $E\left[\frac{\partial l_t}{\partial \theta}(\tau_0)\right] = 0$. Now, we prove $E\left[\frac{\partial l_t}{\partial \alpha}(\tau_0)|\mathcal{G}_t\right] = 0$, we have

$$E\left[\frac{\partial l_t}{\partial \alpha}(\tau_0)|\mathcal{G}_t\right] = -\int_{\mathbb{R}} \frac{\partial f}{\partial \alpha}(x, \psi_0) dx.$$

The function $x \mapsto f(x, \psi_0)$ is in $\mathcal{L}^1$, therefore

$$\frac{\partial \varphi_{\psi_0}(u)}{\partial \alpha} = -\int_{\mathbb{R}} e^{iux} \frac{\partial f}{\partial \alpha}(x, \psi_0) dx, \quad \text{and} \quad \frac{\partial \varphi_{\psi_0}(0)}{\partial \alpha} = E\left[\frac{\partial l_t}{\partial \alpha}(\tau_0)\right].$$

We have, $\forall \alpha,\ \varphi_{\alpha,\beta_0,\mu_0,\gamma_0}(0) = 0$, therefore $\frac{\partial \varphi_{\psi_0}}{\partial \alpha}(0) = 0$. Using the same method for $\frac{\partial l_t}{\partial \beta}$, $\frac{\partial l_t}{\partial \mu}$ and $\frac{\partial l_t}{\partial \gamma}$ we obtain $E\left[\nu_{nt} | \mathcal{G}_t\right] = 0$.

We now prove that the covariance matrix of the vector of derivatives of l_t is finite, we have

$$E\left[\frac{\partial l_t}{\partial \theta}(\tau_0) \frac{\partial l_t}{\partial \theta'}(\tau_0)\right] = \frac{1}{4} E\left[Z_t^2(\tau_0)\right] E\left[\phi_t(\theta_0)\phi_t'(\theta_0)\right],$$

with $E\left[Z_t^2(\tau_0)\right] = 1 + 2\int_{\mathbb{R}} x f'(x, \psi_0) dx + \int_{\mathbb{R}} x^2 \frac{f'^2(x,\psi_0)}{f(x,\psi_0)} dx$. From the asymptotic expansions in (3.2.6)-(3.2.10), we obtain $E\left[Z_t^2(\tau_0)\right] < +\infty$, then using Assumption **A6**, we have $V\left[\frac{\partial l_t}{\partial \theta}(\tau_0)\right]$ is finite.

Using the asymptotic expansion again, we have when $x \to \pm\infty$, for $\psi \in A \times B \times C$

$$\frac{\partial \log f}{\partial \alpha}(x, \psi) \sim K \log|x|,\ \frac{\partial \log f}{\partial \beta}(x, \psi) \sim K,\ \frac{\partial \log f}{\partial \mu}(x, \psi) \sim K x^{-1}, \tag{3.7.7}$$

therefore, for $(i, j) \in \{1, \cdots, 3\}$, we have $E\left[\frac{\partial l_t}{\partial \psi_i} \frac{\partial l_t}{\partial \psi_j}\right]$ is finite. The very same reasoning applies for $E\left[\frac{\partial l_t}{\partial \psi} \frac{\partial l_t}{\partial \theta}\right]$ and we have $V\left[\frac{\partial l_t}{\partial \tau}\right] < +\infty$.

We now show that this matrix is positive-definite. Suppose that we have $(u', v')' \in \mathbb{R}^{m+3}$ such that $(u', v').\frac{\partial l_t}{\partial \tau}(\tau_0) = 0$. We have

$$\left(1 + \eta_t \frac{f'(\eta_t, \psi)}{f(\eta_t, \psi)}\right) \frac{1}{2} u' \phi_t(\theta_0) = v' \frac{\partial \log f}{\partial \psi}(\eta_t, \psi). \tag{3.7.8}$$

Now, we have $\left(1 + \eta_t \frac{f'(\eta_t, \psi)}{f(\eta_t, \psi)}\right) \in \mathcal{G}_{t+1}$, $v' \frac{\partial \log f}{\partial \psi}(\eta_t, \psi) \in \mathcal{G}_{t+1}$ and $\frac{1}{2} u' \phi_t(\theta_0) \in \mathcal{G}_t$. Therefore, we have $\frac{1}{2} u' \phi_t(\theta_0) = Q$ and

$$Q\left(1 + \eta_t \frac{f'(\eta_t, \psi)}{f(\eta_t, \psi)}\right) = v_1 \frac{\partial \log f}{\partial \alpha}(\eta_t, \alpha, \beta) + v_2 \frac{\partial \log f}{\partial \beta}(\eta_t, \alpha, \beta) + v_3 \frac{\partial \log f}{\partial \mu},$$

with $v = (v_1, v_2, v_3)$. We have when $x \to +\infty$, $x \frac{f'(x,\alpha,\beta)}{f(x,\alpha,\beta)} \sim K$, then with Equation (3.7.7) and letting $x \to \infty$, we obtain $v_1 = 0$ and $\forall x \in \mathbb{R}$

$$Q\left(f(x,\psi) + xf'(x,\psi)\right) = v_2 \frac{\partial f}{\partial \beta}(x,\psi) + v_3 \frac{\partial f}{\partial \mu}(x,\psi).$$

Now multiplying both sides of the previous equation by e^{itx} and integrating on $\mathbb{R}$, we recognize the characteristic function of a stable distribution or its derivatives and obtain for $t \in \mathbb{R}$,

$$-Qt\frac{\partial \varphi_\psi(t)}{\partial t} = v_2 \frac{\partial \varphi_\psi(t)}{\partial \beta} + v_3 \frac{\partial \varphi_\psi(t)}{\partial \mu}.$$

Then, for $t > 0$, we obtain

$$Q\left(\alpha t^\alpha - i\beta \tan\frac{\alpha\pi}{2}(\alpha t^\alpha - t)\right) = v_2 i \tan\frac{\alpha\pi}{2}(t^\alpha - t) + v_3 it.$$

Therefore, we have $Q = 0$ and then $v_2 = v_3 = 0$. Finally, with Assumption **A8** we obtain $u = 0_m$.

We have, if $(u', v')\ \frac{\partial l_t}{\partial \tau}(\tau_0) = 0$ then $(u', v')' = 0_{m+3}$ and the matrix $E\left[\frac{\partial l_t}{\partial \tau}(\tau_0)\frac{\partial l_t}{\partial \tau'}(\tau_0)\right]$ is positive. Finally, using the central limit theorem for martingale differences and the Wold-Cramér Lemma we obtain (3.7.4) with $\Lambda = V\left[\frac{\partial l_t}{\partial \tau}\right]$. □

Lemma 3.7.2. *Under the assumptions of Theorem 3.3.1, for any compact subset Γ^* in the interior of Γ, we have*

$$E\left[\sup_{\tau \in \Gamma^*}\left|\frac{\partial^3 l_t}{\partial \tau_i \tau_j \tau_k}(\tau)\right|\right] < +\infty. \tag{3.7.9}$$

Proof. For ease of notation, the next equations are written without their argument $\tau = (\theta, \psi) \in \Gamma$.

We have

$$\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j} = \frac{1}{2}\left(\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\theta_j} - \phi_{t,i}\phi_{t,j}\right)\left(1+\eta_t(\theta)\frac{\partial\log f}{\partial x}\right) \tag{3.7.10}$$
$$-\frac{1}{4}\phi_{t,i}\phi_{t,j}\eta_t(\theta)\left(\frac{\partial\log f}{\partial x}+\eta_t(\theta)\frac{\partial^2\log f}{\partial x^2}\right)$$

$$\frac{\partial^2 l_t}{\partial\theta_i\partial\psi} = \frac{1}{2}\phi_{t,i}\eta_t(\theta)\frac{\partial^2\log f}{\partial x\partial\psi} \tag{3.7.11}$$

$$\frac{\partial^3 l_t}{\partial\theta_i\partial\theta_j\partial\theta_k} = -\left(\phi_{t,i}\phi_{t,j,k}+\phi_{t,j}\phi_{t,i,k}+\phi_{t,k}\phi_{t,i,j}\right)\left(\frac{1}{2}+\frac{3}{4}\eta_t(\theta)\frac{\partial\log f}{\partial x}-\frac{1}{4}\eta_t^2(\theta)\frac{\partial^2\log f}{\partial x^2}\right)$$
$$+\phi_{t,i}\phi_{t,j}\phi_{t,k}\left(1+\frac{15}{8}\eta_t(\theta)\frac{\partial\log f}{\partial x}+\frac{9}{8}\eta_t^2(\theta)\frac{\partial^2\log f}{\partial x^2}+\frac{1}{4}\eta_t^3(\theta)\frac{\partial^3\log f}{\partial x^3}\right)$$
$$+\frac{1}{2}\phi_{t,i,j,k}\left(1+\eta_t(\theta)\frac{\partial\log f}{\partial x}\right),$$

We show that

$$E\sup_{\tau\in\Gamma}\left|\eta_t(\theta)\frac{\partial\log f}{\partial x}(\eta_t(\theta),\psi)\right| < \infty. \tag{3.7.12}$$

For any $\psi\in A\times B\times C$, the function $x\mapsto\left|x\frac{\partial\log f}{\partial x}(x,\psi)\right|$ is bounded. Then the function $\psi\mapsto\sup_x\left|x\frac{\partial\log f}{\partial x}(x,\psi)\right|$ is continuous, thus since $A\times B\times C$ is a compact set, we obtain (3.7.12).

Using the same method for $\eta_t^2(\theta)\frac{\partial^2\log f}{\partial x^2}(\eta_t(\theta),\psi)$ and $\eta_t^3(\theta)\frac{\partial^3\log f}{\partial x^3}(\eta_t(\theta),\psi)$ and using Assumption **A6** we obtain that for any compact subset Γ^* in the interior of Γ

$$E\left[\sup_{\tau\in V}\left|\frac{\partial^3 l_t}{\partial\theta_i\theta_j\theta_k}(\tau)\right|\right] < +\infty.$$

With the same reasoning and other calculations, we obtain Equation (3.7.9). □

Lemma 3.7.3. *Under the assumptions of Theorem 3.3.1, we have when $n\to+\infty$*

$$\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\frac{\partial l_t}{\partial\tau}(\tau_0)-\frac{\partial\tilde{l}_t}{\partial\tau}(\tau_0)\right\}\right\|\to 0 \tag{3.7.13}$$

$$\sup_{\tau\in\Gamma}\left\|\frac{1}{n}\sum_{t=1}^n\left\{\frac{\partial^2 l_t}{\partial\tau\partial\tau'}(\tau)-\frac{\partial^2\tilde{l}_t}{\partial\tau\tau'}(\tau)\right\}\right\|\to 0 \tag{3.7.14}$$

Proof. We have

$$\begin{aligned}\frac{\partial l_t}{\partial \theta}(\tau_0) - \frac{\partial \tilde{l}_t}{\partial \theta}(\tau_0) &= \frac{1}{2}\tilde{\phi}_t \tilde{Z}_t - \frac{1}{2}\phi_t Z_t \\ &= \frac{1}{2}\left\{\tilde{\phi}_t\left(\tilde{Z}_t - Z_t\right) + Z_t\left(\tilde{\phi}_t - \phi_t\right)\right\}. \qquad (3.7.15)\end{aligned}$$

For $\psi \in A\times B\times C$, define the function h as $h(x) = 1+x\frac{\partial \log f}{\partial x}(x,\psi)$, we have $\tilde{Z}_t - Z_t = h(\tilde{\eta}_t) - h(\eta_t)$. When $x \to \pm\infty$, we have $h'(x) = O(x^{-1})$, therefore with the mean value theorem, we have

$$\left|\tilde{Z}_t - Z_t\right| < K|\tilde{\eta}_t - \eta_t| < K|\epsilon_t|\left|\tilde{\sigma}_t^2 - \sigma_t^2\right|$$

On the other side, concerning the second term in (3.7.15), we have

$$\begin{aligned}\left|\tilde{\phi}_t - \phi_t\right| &< \frac{1}{\tilde{\sigma}_t^2}\left|\frac{\partial \tilde{\sigma}_t^2}{\partial \theta} - \frac{\partial \sigma_t^2}{\partial \theta}\right| + \frac{\partial \sigma_t^2}{\partial \theta}\left|\frac{1}{\tilde{\sigma}_t^2} - \frac{1}{\sigma_t^2}\right| \\ &< K\left|\frac{\partial \tilde{\sigma}_t^2}{\partial \theta} - \frac{\partial \sigma_t^2}{\partial \theta}\right| + \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\left|\frac{\sigma_t^2}{\tilde{\sigma}_t^2} - 1\right|.\end{aligned}$$

Concerning the derivatives relative to the stable parameter ψ, we have with the mean value theorem.

$$\left|\frac{\partial \log f}{\partial \alpha}(\tilde{\eta}_t,\psi) - \frac{\partial \log f}{\partial \alpha}(\eta_t,\psi)\right| < \sup_{x\in\mathbb{R}}\left|\frac{\partial^2 \log f}{\partial x \partial \alpha}(x,\psi)\right| |\tilde{\eta}_t - \eta_t|.$$

The derivative of $\log f$ relative to α and x is bounded and we have

$$|\tilde{\eta}_t - \eta_t| = \left|\frac{\epsilon_t}{\tilde{\sigma}_t} - \frac{\epsilon_t}{\sigma_t}\right| < K|\epsilon_t|\left|\tilde{\sigma}_t^2 - \sigma_t^2\right|.$$

We can apply the same method for the derivatives relative to β and μ. Therefore, using Assumption **A7**, the Markov inequality, the Borel-Cantelli lemma and the Cesàro Lemma, we easily obtain (3.7.13).

Using (3.7.10), (3.7.11), the second part of Assumption **A7** and the same techniques as before, we obtain (3.7.14). □

Proof of Theorem 3.3.1. From the definition of the ML estimator τ_n, we have $\frac{\partial \tilde{I}_n}{\partial \tau}(\tau_n) = 0$, writing a Taylor expansion, we obtain

$$0 = \frac{\partial \tilde{I}_n}{\partial \tau}(\tau_n) = \frac{\partial \tilde{I}_n}{\partial \tau}(\tau_0) + \frac{\partial^2 \tilde{I}_n}{\partial \tau \partial \tau'}(\tau^*)(\tau_n - \tau_0),$$

where τ^* is between τ_0 and τ_n. Using another Taylor expansion, Lemma 3.7.2, the almost sure convergence of τ_n to τ_0 and the ergodic theorem, we obtain

$$\frac{\partial^2 I_n}{\partial \tau \partial \tau'}(\tau^*) \to E\left[\frac{\partial^2 l_t}{\partial \tau \partial \tau'}(\tau_0)\right], \quad \text{a.s.}$$

Then, using Equation (3.7.5), (3.7.6), (3.7.10) and (3.7.11), we obtain

$$E\left[\frac{\partial^2 l_t}{\partial \tau \partial \tau'}(\tau_0)\right] = E\left[\frac{\partial l_t}{\partial \tau}(\tau_0)\frac{\partial l_t}{\partial \tau'}(\tau_0)\right].$$

By Lemmas 3.7.1 and 3.7.3 we can conclude and obtain (3.3.3). □

3.7.3 Proof of the consistency in Theorem 3.4.1

Let I_n (resp. l_{nt}) be the equivalent of $\tilde{I}_n$ (resp. $\tilde{l}_{nt}$) when an infinite past is known.

$$I_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} l_{nt}(\tau), \text{ with } \quad l_{nt}(\tau) = \frac{1}{2}\log\sigma_{nt}^2(\theta) - \log f\left(\frac{\epsilon_{nt}}{\sigma_{nt}(\theta)}, \psi\right).$$

We also need to define the equivalent of these quantities when the processus (η_{nt}) is replaced by its limit in distribution (η_t). If (ϵ_t) is the stationary ergodic solution of Model (3.4.3), we define $\sigma_t^2(\theta) = \frac{\omega}{\mathcal{B}_\theta(1)} + \mathcal{B}_\theta^{-1}(L)\mathcal{A}_\theta(L)\epsilon_t^2$ and $l_t(\tau) = \frac{1}{2}\log\sigma_t^2(\theta) - \log f\left(\frac{\epsilon_t}{\sigma_t(\theta)}, \psi\right)$.

Assumption **B3** can only be used for quantities which depend on a finite number of $(\eta_{nt})_t$. Therefore we introduce $\sigma_{nt}^{2(m)}(\theta)$, a truncated version of $\sigma_{nt}^2(\theta)$. For that, we give a vector representation of the GARCH(p,q) model as in Bougerol and Picard (1992),

$$\underline{z}_{nt} = \underline{b}_{nt} + A_{nt}\underline{z}_{nt-1},$$

where

$$\underline{b}_{nt} = \begin{pmatrix} \omega_0\eta_{nt}^2 \\ 0 \\ \vdots \\ \omega_0 \\ 0 \\ \vdots \end{pmatrix} \in \mathbb{R}^{p+q}, \ \underline{z}_{nt} = \begin{pmatrix} \epsilon_{nt}^2 \\ \vdots \\ \epsilon_{nt-q+1}^2 \\ \sigma_{nt}^2 \\ \vdots \\ \sigma_{nt-p+1}^2 \end{pmatrix} \in \mathbb{R}^{p+q},$$

and

$$A_{nt} = \begin{pmatrix} a_{01}\eta_{nt}^2 & & \cdots & & a_{0q}\eta_{nt}^2 & b_{01}\eta_{nt}^2 & & \cdots & & b_{0p}\eta_{nt}^2 \\ 1 & 0 & \cdots & & 0 & 0 & & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 & 0 & & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \cdots & 1 & 0 & 0 & & \cdots & 0 & 0 \\ a_{01} & & \cdots & & a_{0q} & b_{01} & & \cdots & & b_{0p} \\ 0 & & \cdots & & 0 & 1 & 0 & \cdots & & 0 \\ 0 & & \cdots & & 0 & 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \cdots \\ 0 & & \cdots & 0 & 0 & 0 & & \cdots & 1 & 0 \end{pmatrix}.$$

We also define $\underline{z}_t$, $\underline{b}_t$ and A_t, the counterparts of $\underline{z}_{nt}$, $\underline{b}_{nt}$ and A_{nt} when η_{nt} is replaced by the iid sequence (η_t) defined in (3.4.3). Note that γ_n is the top Lyapunov exponent associated to the sequence $(A_{nt})_{t\in\mathbb{Z}}$. Now, we prove that Assumption **B2** implies that γ_n is inferior to zero.

With Lemma 3.7.4 below, we obtain for any $n \in \mathbb{N}$,

$$\underline{z}_{nt} = \underline{b}_{nt} + \sum_{k=1}^{+\infty} A_{nt}A_{nt-1}\dots A_{nt-k+1}\underline{b}_{nt-k}. \tag{3.7.16}$$

We define the truncated version of $\underline{z}_{nt}$. For any $m \in \mathbb{N}$,

$$\underline{z}_{nt}^{(m)} = \underline{b}_{nt} + \sum_{k=1}^{m} A_{nt}A_{nt-1}\dots A_{nt-k+1}\underline{b}_{nt-k},\ \forall t \in \mathbb{Z},\ \forall n \in \mathbb{N}. \tag{3.7.17}$$

In particular, if $X(k)$ is the k^{th} element of the vector X, we define a truncated version of σ_{nt}^2, that is

$$\underline{z}_{nt}^{(m)}(q+1) = \underline{b}_{nt}(q+1) + \left(\sum_{k=1}^{m} A_{nt}A_{nt-1}\dots A_{nt-k+1}\underline{b}_{nt-k}\right)(q+1). \tag{3.7.18}$$

The quantity $\underline{z}_{nt}^{(m)}(q+1)$ depends only on $\{\eta_{nt-1},\dots,\eta_{nt-m}\}$.

Then we define $\sigma_{nt}^{2(m)}(\theta)$ for any $\theta \in \Theta$. For that, we introduce another vector representation of the model,

$$\underline{\sigma}_{nt}^2(\theta) = \underline{c}_{nt}(\theta) + B\underline{\sigma}_{nt-1}^2(\theta),$$

where

$$\underline{\sigma}_{nt}^2(\theta) = \begin{pmatrix} \sigma_{nt}^2(\theta) \\ \sigma_{nt-1}^2(\theta) \\ \vdots \\ \sigma_{nt-p+1}^2(\theta) \end{pmatrix},\ \underline{c}_{nt}(\theta) = \begin{pmatrix} \omega + \sum_{i=1}^q a_i\epsilon_{nt-i}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix},\ B = \begin{pmatrix} b_1 & b_2 & \cdots & b_p \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & 1 & 0 \end{pmatrix}.$$

From Assumption **B2**, we have $\sup_{\theta\in\Theta} \rho(B) < 1$, where $\rho(B)$ is the spectral radius of the matrix B and thus for any $\theta \in \Theta$,

$$\underline{\sigma}^2_{nt}(\theta) = \sum_{k=0}^{+\infty} B^k \underline{c}_{nt-k}(\theta). \tag{3.7.19}$$

We define for $m \in \mathbb{N}$,

$$\sigma^{2(m)}_{nt}(\theta) = \sum_{k=0}^{m} B^k(1,1)\underline{c}^{(m)}_{nt-k}(\theta)(1), \text{ with } \underline{c}^{(m)}_{nt}(\theta)(1) = \omega + \sum_{i=1}^{q} a_i \underline{z}^{(m)}_{nt-i}(q+1)\eta^2_{nt-i}. \tag{3.7.20}$$

As for $\underline{z}^{(m)}_{nt}(q+1)$, the quantity $\sigma^{2(m)}_{nt}(\theta)$ depends on a finite number of $\eta_{nt'}$, but since every $\underline{z}^{(m)}_{nt-i}(q+1)$ depends on several $\eta_{nt'}$, $\sigma^{2(m)}_{nt}(\theta)$ depends on more than m variables $\eta_{nt'}$. To be precise, $\sigma^{2(m)}_{nt}(\theta)$ depends on $\{\eta_{nt-1}, \dots, \eta_{nt-2m+q}\}$. Define also $l^{(m)}_{nt}(\tau) = \frac{1}{2}\log \sigma^{2(m)}_{nt}(\theta) - \log f\left(\frac{\epsilon_{nt}}{\sigma^{2(m)}_{nt}(\theta)}, \psi\right)$.

Lemma 3.7.4. *Under the assumptions of Theorem 3.4.1, there exists $N \in \mathbb{N}$ such that*

$$\forall n \geq N, \ \gamma_n < 0. \tag{3.7.21}$$

Besides, there exist $k_0 \in \mathbb{N}$ and $N \in \mathbb{N}$ such that

$$\chi' = \sup_{n\geq N} E\left[\|A_{nk_0}A_{nk_0-1}\cdots A_{n1}\|^s\right] < 1. \tag{3.7.22}$$

Proof of Lemma 3.7.4. With Assumption **B2**, using the norm $\|A\| = \sum |a_{ij}|$, which is a multiplicative norm and with Lemma 2.3 in Francq and Zakoian (2010), we have the existence of $k_0 \in \mathbb{N}$ and of $s > 0$ such that

$$\chi = E\left[\|A_{k_0}A_{k_0-1}\cdots A_1\|^s\right] < 1.$$

Now for $n \in \mathbb{N}$, writing $A_{nt} = A(\eta_{nt})$ to emphasize the fact that A_{nt} only depends on η_{nt}, we have for $s > 0$

$$E\left[\|A_{nk_0}A_{nk_0-1}\cdots A_{1n}\|^s\right] = \int_{\mathbb{R}^{k_0}} \|A(x_1)\cdots A(x_{k_0})\|^s f_n(x_1)\cdots f_n(x_{k_0})dx_1\cdots dx_{k_0}.$$

For $\varepsilon > 0$, with Assumption **B3**, we have the existence of $N \in \mathbb{N}$, such that

$$\forall n \geq N,\ \forall x \in \mathbb{R},\ f_n(x) \leq G(x),\ \text{where } G(x) = f(x) + \frac{\varepsilon}{(1+|x|)^{\delta}}.$$

Then, the function A is such that $\forall x \in \mathbb{R},\ 0 < \|A(x)\| < Kx^2$ and therefore

$$\|A(x_1)\cdots A(x_{k_0})\|^s f_n(x_1)\cdots f_n(x_{k_0}) \leq K \prod_{i=1}^{k_0} |x_i|^{2s} G(x_i),$$

Using the asymptotic expansion (3.2.6) and choosing $0 < s < \min\left(\frac{\delta-1}{2}, \frac{\alpha}{2}\right)$, we obtain

$$\int_{\mathbb{R}^{k_0}} K \prod_{i=1}^{k_0} |x_i|^{2s} G(x_i) \prod_{i=1}^{k_0} dx_i < +\infty,$$

Therefore, since f_n simply converges to $f(., \psi_0)$, using the dominated convergence theorem, we obtain

$$\lim_{n\to+\infty} E\left[\|A_{nk_0}A_{nk_0-1}\cdots A_{1n}\|^s\right] = E\left[\|A_{k_0}A_{k_0-1}\cdots A_1\|^s\right] = \chi < 1.$$

Therefore for $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for $n \geq N$, we have

$$E\left[\|A_{nk_0}A_{nk_0-1}\cdots A_{n1}\|^s\right] < 1 - \varepsilon,$$

and thus $\chi' = \sup_{n\geq N} E\left[\|A_{nk_0}A_{nk_0-1}\cdots A_{n1}\|^s\right] < 1$ and we obtain (3.7.22).

Then using Lemma 2.3 from Francq and Zakoian (2010) again, we obtain (3.7.21). □

Lemma 3.7.5. *Under the assumptions of Theorem 3.4.1, there exists $s > 0$ such that,*

$$\sup_{n\in\mathbb{N}} E\,|\epsilon_{nt}|^{2s} < +\infty, \quad \textit{and} \quad \sup_{n\in\mathbb{N}} E\sigma_{nt}^{2s} < +\infty. \tag{3.7.23}$$

Proof. For $n \geq N$, using the inequality $(x+y)^s \leq x^s + y^s$ for $x, y > 0$ and $s < 1$, Equation (3.7.16), the fact that the norm is multiplicative, the independence of the processus $(\eta_{nt})_t$ and Lemma 3.7.4, we obtain

$$E\,\|\underline{z}_{nt}\|^s \leq \|E\underline{b}_{n1}\|^s \left\{1 + \sum_{k=0}^{+\infty} \chi'^k \sum_{i=1}^{k_0} \left\{E\,\|A_{n1}\|^s\right\}^i\right\}. \tag{3.7.24}$$

Now, we prove that there exists $s > 0$ such that $\sup_{n\in\mathbb{N}} E|\eta_{nt}|^{2s} < +\infty$. In view of Assumption **B3**, we obtain that $E|\eta_{nt}|^{2s}$ converges toward $E|\eta_t|^{2s} < +\infty$, for $s < \delta/2$, with the dominated convergence theorem again and also the fact that for any $n \in \mathbb{N}$, we have $E|\eta_{nt}|^{2s} < +\infty$ for a small enough $s > 0$. Therefore with (3.7.24), we have

$$\sup_{n\in\mathbb{N}} E\left\|\underline{z}_{nt}\right\|^s < +\infty.$$

Now, since for any $n \in \mathbb{N}$ and any $t \in \mathbb{Z}$, we have $\sigma_{nt}^2 \leq \|\underline{z}_{nt}\|$ and $\epsilon_{nt}^2 \leq \|\underline{z}_{nt}\|$, we obtain (3.7.23). □

Lemma 3.7.6. *Under the assumptions of Theorem 3.4.1, there exists $s > 0$ such that,*

$$\sup_{n\in\mathbb{N}}\sup_{\theta\in\Theta} E\left|\sigma_{nt}^{2s}(\theta) - \sigma_{nt}^{2(m)s}(\theta)\right| < K\rho^m. \tag{3.7.25}$$

Proof. We first prove that there exists $s > 0$ such that

$$\sup_{n\in\mathbb{N}} E\left|\sigma_{nt}^{2s}(\theta_0) - \underline{z}_{nt}^{(m)s}(q+1)\right| < K\rho^m. \tag{3.7.26}$$

For $m \geq k_0^2$, let $\lfloor m/k_0 \rfloor$ be the floor function of m/k_0 (k_0 being defined as in Lemma 3.7.4), we have

$$\left\|\underline{z}_{nt} - \underline{z}_{nt}^{(\lfloor m/k_0\rfloor)}\right\| = \sum_{k=\lfloor m/k_0\rfloor+1}^{+\infty} \|A_{nt}\cdots A_{nt-k+1}\| \left\|\underline{b}_{nt-k}\right\|.$$

The constant s can be taken such that $s < 1$ and we obtain, using the inequality $(x+y)^s \leq x^s + y^s$ for $x, y > 0$ and using the independence of the processus $(\eta_{nt})_t$,

$$\begin{aligned}
\sup_{n\geq N} E\left[\left\|\underline{z}_{nt} - \underline{z}_{nt}^{(\lfloor m/k_0\rfloor)}\right\|^s\right] &\leq \sum_{k=\lfloor m/k_0\rfloor+1}^{+\infty} \sup_{n\geq N} E\left[\|A_{nt}\cdots A_{nt-k+1}\|^s\right] \sup_{n\geq N} E\left[\left\|\underline{b}_{nt-k}\right\|^s\right] \\
&\leq \sup_{n\geq N} E\left[\|\underline{b}_{n1}\|^s\right] \sum_{k=\lfloor m/k_0\rfloor+1}^{+\infty} \chi'^k \sum_{i=1}^{k_0} \left\{\sup_{n\geq N} E\left\|A_{n1}\right\|^s\right\}^i \\
&\leq K\rho^m,
\end{aligned}$$

defining $N \in \mathbb{N}$ and using similar arguments as in the proof of Lemma 3.7.5. With exactly the same arguments, we obtain for any $n \in \mathbb{N}$ the existence of $K_n > 0$ and $\rho_n < 1$ such that

$$E\left[\left\|\underline{z}_{nt} - \underline{z}_{nt}^{(\lfloor m/k_0 \rfloor)}\right\|^s\right] \leq K_n \rho_n^m.$$

Thus, there exist $K > 0$ and $\rho < 1$ such that

$$\sup_{n\in\mathbb{N}} E\left[\left\|\underline{z}_{nt} - \underline{z}_{nt}^{(\lfloor m/k_0 \rfloor)}\right\|^s\right] \leq K\rho^m.$$

Then we use the inequality $\left|\sigma_{nt}^{2s} - \underline{z}_{nt}^{(m)s}(q+1)\right| \leq \left|\sigma_{nt}^{2} - \underline{z}_{nt}^{(m)}(q+1)\right|^s \leq \left\|\underline{z}_{nt} - \underline{z}_{nt}^{(m)}\right\|^s$ and obtain Equation (3.7.26). We also obtain

$$\sup_{n\in\mathbb{N}} E\left[\left|\sigma_{nt}^{2} - \underline{z}_{nt}^{(m)}(q+1)\right|^s\right] \leq K\rho^m. \tag{3.7.27}$$

We now prove the inequality (3.7.25). We remark that for any $m \in \mathbb{N}$ and for any $\theta \in \Theta$, we have $\sigma_{nt}^{2(m)}(\theta) \leq \sigma_{nt}^2(\theta)$. Then, we have

$$\begin{aligned}\sup_{n\in\mathbb{N}} \sup_{\theta\in\Theta} E\left[\left|\sigma_{nt}^{2s}(\theta) - \sigma_{nt}^{2(m)s}(\theta)\right|\right] &\leq \sum_{k=m+1}^{+\infty} \sup_{\theta\in\Theta} B^k(1,1)^s \sup_{n\in\mathbb{N}} E\left[\sup_{\theta\in\Theta} \underline{c}_{nt-k}^s(\theta)(1)\right] \\ &+ \sum_{k=0}^{m} \sup_{\theta\in\Theta} B^k(1,1)^s \sum_{i=1}^{q} \sup_{\theta\in\Theta} a_i^s \sup_{n\in\mathbb{N}} E\eta_{nt-i-k}^{2s} \sup_{n\in\mathbb{N}} E\left[\left(\sigma_{nt-i-k}^{2} - \underline{z}_{nt-i-k}^{(m)}(q+1)\right)^s\right]\end{aligned} \tag{3.7.28}$$

With the second part of Assumption **B2**, we have $\sup_{\theta\in\Theta} \rho(B) < 1$. Then, using Lemma 3.7.5, we obtain

$$\sum_{k=m+1}^{+\infty} \sup_{\theta\in\Theta} B^k(1,1)^s \sup_{n\in\mathbb{N}} E\left[\sup_{\theta\in\Theta} \underline{c}_{nt-k}^s(\theta)(1)\right] \leq K\rho^m.$$

Now for the second part of (3.7.28), using Equation (3.7.27) and the fact that $\rho(B) < 1$, we obtain

$$\sum_{k=0}^{m} \sup_{\theta\in\Theta} B^k(1,1)^s \sum_{i=1}^{q} \sup_{\theta\in\Theta} a_i^s \sup_{n\in\mathbb{N}} E\eta_{nt-i-k}^{2s} \sup_{n\in\mathbb{N}} E\left[\left|\sigma_{nt-i-k}^{2} - \underline{z}_{nt-i-k}^{2(m)}(q+1)\right|^s\right] \leq K\rho^m.$$

Finally, having treated the two terms of the right hand of (3.7.28), we obtain (3.7.25). □

Lemma 3.7.7. *Under the assumptions of Theorem 3.4.1, we have for any $d \in \mathbb{N}$ and for any* *subset* $V \subset \Gamma$

$$E\left[\left|\inf_{\tau\in V} l_t(\tau)\right|^d\right] < +\infty, \tag{3.7.29}$$

$$E\left[\left(\inf_{\tau\in V} l_{nt}(\tau)\right)^d\right] \underset{n\to+\infty}{\longrightarrow} E\left[\left(\inf_{\tau\in V} l_t(\tau)\right)^d\right] \tag{3.7.30}$$

$$E\left[\left|\inf_{\tau\in V} l_{nt}(\tau)\right|^d\right] \underset{n\to+\infty}{\longrightarrow} E\left[\left|\inf_{\tau\in V} l_t(\tau)\right|^d\right]. \tag{3.7.31}$$

Proof. We prove (3.7.30) in the case $d = 1$. The other cases and (3.7.31) can be obtained with similar arguments. We will prove the following intermediate results. For any subset $V \subset \Gamma$

(i) $\sup_{n\in\mathbb{N}} E\left|\inf_{\tau\in V} l_{nt}(\tau) - \inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right| < K\rho^m$.

(ii) $E\left|\inf_{\tau\in V} l_t(\tau) - \inf_{\tau\in V} l_t^{(m)}(\tau)\right| < K\rho^m$.

(iii) For any $m > 0$, $E\inf_{\tau\in V} l_{nt}^{(m)}(\tau) \to E\inf_{\tau\in V} l_t^{(m)}(\tau)$, when $n \to +\infty$.

We have for any $\theta \in \Theta$, $\sigma_{nt}^2(\theta) \geq \omega$. Since Θ is a compact set, there exists $\underline{\omega} > 0$ such that, $\forall\theta \in \Theta$, $\forall t \in \mathbb{Z}$, $\forall n \in \mathbb{N}$, $\sigma_{nt}^2(\theta) \geq \underline{\omega}$ and $\forall\theta \in \Theta$, $\forall t \in \mathbb{Z}$, $\forall n \in \mathbb{N}$, $\sigma_{nt}^{2(m)}(\theta) \geq \underline{\omega}$. From Lemma 3.7.6 and using the mean value theorem we have

$$\sup_{n\in\mathbb{N}}\sup_{\theta\in\Theta} E\left|\log\sigma_{nt}^2(\theta) - \log\sigma_{nt}^{2(m)}(\theta)\right| \leq K\sup_{n\in\mathbb{N}}\sup_{\theta\in\Theta} E\left|\sigma_{nt}^{2s}(\theta) - \sigma_{nt}^{2(m)s}(\theta)\right| < K\rho^m. \tag{3.7.32}$$

For $\theta \in \Theta$, let $a_{nt}(\theta) = \frac{\sigma_{nt}(\theta_0)}{\sigma_{nt}(\theta)}$ and let $a_{nt}^{(m)}(\theta) = \sqrt{\frac{\sigma_{nt}^{2(m)}(\theta_0)}{\sigma_{nt}^{2(m)}(\theta)}}$. We have for $s' > 0$

$$\left|a_{nt}^{2s'}(\theta) - a_{nt}^{(m)2s'}(\theta)\right| \leq \sigma_{nt}^{2s'}(\theta_0)\left|\frac{1}{\sigma_{nt}^{2s'}(\theta)} - \frac{1}{\sigma_{nt}^{2(m)s'}(\theta)}\right| + \frac{1}{\sigma_{nt}^{2(m)s'}(\theta)}\left|\sigma_{nt}^{2s'}(\theta_0) - \sigma_{nt}^{2(m)s'}(\theta_0)\right|,$$

Setting $s' = s/2$ and using the Cauchy-Schwarz inequality and the results of Lemmas 3.7.5 and 3.7.6, we obtain

$$\sup_{n\in\mathbb{N}}\sup_{\theta\in\Theta} E\left|a_{nt}^{s'}(\theta) - a_{nt}^{(m)s'}(\theta)\right| < K\rho^m.$$

Then, using the independence between $\sigma_{nt}^2(\theta)$ (or $\sigma_{nt}^{2(m)}$) and η_{nt} and Assumption **B3**, we obtain for any $\theta \in \Theta$

$$\sup_{n\in\mathbb{N}}\sup_{\theta\in\Theta} E\left[|\eta_{nt}|^{s'}\left|a_{nt}^{s'}(\theta) - a_{nt}^{(m)s'}(\theta)\right|\right] < K\rho^m.$$

Defining the function $F_\psi(x) = \log f(x^{1/s'}, \psi)$, we have, if $\eta_{nt} > 0$

$$\left|\log f\left(a_{nt}(\theta)\eta_{nt}, \psi\right) - \log f\left(a_{nt}^{(m)}(\theta)\eta_{nt}, \psi\right)\right| = \left|F_\psi\left(a_{nt}^{s'}(\theta)|\eta_{nt}|^{s'}\right) - F_\psi\left(a_{nt}^{(m)s'}(\theta)|\eta_{nt}|^{s'}\right)\right|.$$

The derivative of F is such that $\frac{\partial F(x)}{\partial x} = x^{1/s'-1}\frac{f'(x^{1/s'},\psi)}{f(x^{1/s'},\psi)}$. We have, when $x \to +\infty$, $\frac{\partial F_\psi}{\partial x} \sim 1/x$. Therefore if we take $s' < 1$ we obtain that $\frac{\partial F_\psi}{\partial x}$ is bounded. Then since Γ is a compact set and since $\psi \mapsto \sup_x \frac{\partial F_\psi}{\partial x}(x)$ is continuous, we obtain $\sup_{\tau\in\Gamma}\sup_x \frac{\partial F_\psi}{\partial x}(x) < +\infty$. Then, with the mean value theorem, we have

$$\sup_{\tau\in\Gamma}\left|\log f\left(a_{nt}(\theta)\eta_{nt}, \psi\right) - \log f\left(a_{nt}^{(m)}(\theta)\eta_{nt}, \psi\right)\right| \leq K|\eta_{nt}|^{s'}\sup_{\theta\in\Theta}\left|a_{nt}^{s'}(\theta) - a_{nt}^{(m)s'}(\theta)\right|,$$

and finally

$$\sup_{n\in\mathbb{N}}\sup_{\tau\in\Gamma} E\left|\log f\left(a_{nt}(\theta)\eta_{nt}, \psi\right) - \log f\left(a_{nt}^{(m)}(\theta)\eta_{nt}, \psi\right)\right| < K\rho^m. \tag{3.7.33}$$

With Equations (3.7.32) and (3.7.33), we obtain

$$\sup_{n\in\mathbb{N}}\sup_{\tau\in\Gamma} E\left|l_{nt}(\tau) - l_{nt}^{(m)}(\tau)\right| < K\rho^m. \tag{3.7.34}$$

Now for $m \in \mathbb{N}$, for $K_1 > 0$ and for $|\rho_1| < 1$, for any $n \in \mathbb{N}$, there exists $\tilde{\tau}_{m,n} \in \Gamma$ such that $l_{nt}\left(\tilde{\tau}_{m,n}\right) - \inf_{\tau\in V} l_{nt}(\tau) < K_1\rho_1^m$ and there exists $\hat{\tau}_{m,n} \in \Gamma$ such that $l_{nt}^{(m)}\left(\hat{\tau}_{m,n}\right) - \inf_{\tau\in V} l_{nt}^{(m)}(\tau) < K_1\rho_1^m$. Now if $l_{nt}\left(\hat{\tau}_{m,n}\right) \leq l_{nt}\left(\tilde{\tau}_{m,n}\right)$, we have

$$\begin{aligned}\left|\inf_{\tau\in V} l_{nt}(\tau) - \inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right| &\leq \left|\inf_{\tau\in V} l_{nt}(\tau) - l_{nt}\left(\hat{\tau}_{m,n}\right)\right| + \left|l_{nt}\left(\hat{\tau}_{m,n}\right) - \inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right| \\ &\leq \left|\inf_{\tau\in V} l_{nt}(\tau) - l_{nt}\left(\tilde{\tau}_{m,n}\right)\right| + K_1\rho_1^m \leq K\rho^m.\end{aligned}$$

Or if $l_{nt}^{(m)}\left(\tilde{\tau}_{m,n}\right) \leq l_{nt}^{(m)}\left(\hat{\tau}_{m,n}\right)$, we have

$$\begin{aligned}\left|\inf_{\tau\in V} l_{nt}(\tau) - \inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right| &\leq \left|\inf_{\tau\in V} l_{nt}(\tau) - l_{nt}\left(\tilde{\tau}_{m,n}\right)\right| + \left|l_{nt}\left(\tilde{\tau}_{m,n}\right) - \inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right| \\ &\leq K_1\rho_1^m + \left|l_{nt}^{(m)}\left(\hat{\tau}_{m,n}\right) - \inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right| \leq K\rho^m.\end{aligned}$$

Now, if $l_{nt}\left(\hat{\tau}_{m,n}\right) > l_{nt}\left(\tilde{\tau}_{m,n}\right)$ and $l_{nt}^{(m)}\left(\tilde{\tau}_{m,n}\right) > l_{nt}^{(m)}\left(\hat{\tau}_{m,n}\right)$, we have

$$\begin{aligned}\left|\inf_{\tau\in V} l_{nt}(\tau) - \inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right| \quad &\leq \quad \left|\inf_{\tau\in V} l_{nt}(\tau) - l_{nt}\left(\tilde{\tau}_{m,n}\right)\right| + \left|l_{nt}\left(\tilde{\tau}_{m,n}\right) - l_{nt}^{(m)}\left(\hat{\tau}_{m,n}\right)\right| \\ &\quad + \left|l_{nt}^{(m)}\left(\hat{\tau}_{m,n}\right) - \inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right|. \qquad (3.7.35)\end{aligned}$$

We have

$$l_{nt}\left(\tilde{\tau}_{m,n}\right) - l_{nt}^{(m)}\left(\tilde{\tau}_{m,n}\right) \leq l_{nt}\left(\tilde{\tau}_{m,n}\right) - l_{nt}^{(m)}\left(\hat{\tau}_{m,n}\right) \leq l_{nt}\left(\hat{\tau}_{m,n}\right) - l_{nt}^{(m)}\left(\hat{\tau}_{m,n}\right),$$

and thus, with (3.7.34) we obtain $E\left|l_{nt}\left(\tilde{\tau}_{m,n}\right) - l_{nt}^{(m)}\left(\hat{\tau}_{m,n}\right)\right| < K\rho^m$. Finally, with Equation (3.7.35) we obtain (i), the step (ii) can be obtained in the exact same way.

[step (iii)] We have, for $m \in \mathbb{N}^*$ and $\tau \in \Gamma$, $l_{nt}^{(m)}(\tau) = \frac{1}{2}\log\sigma_{nt}^{2(m)}(\theta) - \log f\left(a_{nt}^{(m)}(\theta)\eta_{nt}, \psi\right)$. The quantity $\sigma_{nt}^{2(m)}(\theta)$ depends on a finite number of η_{nt}. More precisely $\sigma_{nt}^{2(m)}(\theta)$ is a function of $\{\eta_{nt-k},\ k \in \{1,\ldots,2m+q\}\}$. Now, from (3.7.19) we obtain that the expression of $\sigma_{nt}^{2(m)}(\theta)$ contains only products of powers of $\eta_{nt'}$. Therefore, since Θ is a compact set, there exist $M > 0$ and $(r_1,\ldots,r_{2m+q}) \in \mathbb{N}^{2m+q}$ such that

$$\forall\theta\in\Theta,\ \underline{\omega} \leq \sigma_{nt}^{2(m)}(\theta) \leq K\max(M,\eta_{nt-1}^2)^{r_1}\ldots\max(M,\eta_{nt-2m-q}^2)^{r_{2m+q}}. \qquad (3.7.36)$$

Using the same arguments, we have

$$\forall\theta\in\Theta,\ a_{nt}^{(m)}(\theta) \leq K\max(M,\eta_{nt-1}^2)^{s_1}\ldots\max(M,\eta_{nt-2m+q}^2)^{s_{2m+q}}.$$

Then, with the asymptotic expansion (3.2.6), we have $\forall x\in\mathbb{R},\ f(x,\psi) \geq Kx^{-\alpha-1}$ and $\forall\psi \in A\times B\times C,\ \forall x\in\mathbb{R},\ f(x,\psi) < K$. Therefore, there exist $(s_0,\ldots,s_{2m+q})$ such that

$$\forall\tau\in\Gamma,\ K\max(M,\eta_{nt}^2)^{s_0}\ldots\max(M,\eta_{nt-2m-q}^2)^{s_{2m+q}} \leq f(a_{nt}(\theta)\eta_{nt},\psi) \leq K. \qquad (3.7.37)$$

With (3.7.36) and (3.7.37), we obtain the existence of $M > 0$ and $u_i,\ i \in \{0,\ldots,2m+q\}$ such that

$$\left|\inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right| < K\left(1 + \sum_{i=0}^{2m+q}\left|u_i\log\left\{\max(M,\eta_{nt-i}^2)\right\}\right|\right)$$

Then, we can apply the dominated convergence theorem as we did before and obtain (3.7.29) and (iii).

Now, to obtain (3.7.30), using (i), (ii) and (iii) we have

$$\begin{aligned}\lim_{n\to+\infty} E\left[\inf_{\tau\in V} l_{nt}(\tau)\right] &= \lim_{n\to+\infty}\lim_{m\to+\infty} E\left[\inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right]\\ &= \lim_{m\to+\infty}\lim_{n\to+\infty} E\left[\inf_{\tau\in V} l_{nt}^{(m)}(\tau)\right]\\ &= \lim_{m\to+\infty} E\left[\inf_{\tau\in V} l_{t}^{(m)}(\tau)\right]\\ &= E\left[\inf_{\tau\in V} l_t(\tau)\right].\end{aligned}$$

The limits inversion can be done since the convergence in m is uniform with respect to n. □

Lemma 3.7.8. *Under the assumptions of Theorem 3.4.1, for any subset $V \subset \Gamma$, we have*

$$\frac{1}{n}\sum_{t=1}^{n} \inf_{\tau\in V} l_{nt}(\tau) \to E\left[\inf_{\tau\in V} l_t(\tau)\right], \ \textit{a.s. when } n \to +\infty. \tag{3.7.38}$$

Proof. Let $X_{nt} = (\inf_{\tau\in V} l_{nt}(\tau))^+$ and let $M_n = \frac{1}{n}\sum_{t=1}^{n} X_{nt}$. We also define $S_n = M_{n^2}$ and $m_n = EX_{nt}$. We have $ES_n = m_{n^2}$ and $\mathrm{Var}S_n = \frac{1}{n^2}\sum_{h=0}^{n^2-1}\mathrm{Cov}\,(X_{nt}, X_{nt-h})$. We now prove that there exists $M > 0$ such that for any $n \in \mathbb{N}$ we have

$$\left|\sum_{h=0}^{n^2-1}\mathrm{Cov}\,(X_{nt}, X_{nt-h})\right| < M$$

As in the proof of the previous Lemma, we define $X_{nt}^{(m)} = (\inf_{\tau\in V} l_{nt}^{(m)}(\tau))^+$ and $R_{nt}^{(m)} = X_{nt} - X_{nt}^{(m)}$. With the step (i) of the proof of Lemma 3.7.7, we have for any $m \in \mathbb{N}$

$$\sup_{n\in\mathbb{N}} E\left|R_{nt}^{(m)}\right| < K\rho^m \quad \text{and} \quad \sup_{n\in\mathbb{N}} ER_{nt}^{(m)2} < K\rho^m. \tag{3.7.39}$$

Now we have for $h \in \mathbb{N}$

$$\mathrm{Cov}\,(X_{nt}, X_{nt-h}) = \mathrm{Cov}\left(X_{nt}^{(\lfloor h/2\rfloor)}, X_{nt-h}\right) + \mathrm{Cov}\left(R_{nt}^{(\lfloor h/2\rfloor)}, X_{nt-h}\right).$$

Using (3.7.39), we obtain

$$\sup_{n\in\mathbb{N}} \left|\mathrm{Cov}\left(R_{nt}^{(\lfloor h/2\rfloor)}, X_{nt-h}\right)\right| < K\rho^h. \tag{3.7.40}$$

Now, we know that, for any $\tau \in \Gamma$, $l_{nt}(\tau)$ can be written as a measurable function of $(\epsilon_{nt'})_{t'\leq t}$. Therefore X_{nt-h} is also a measurable function of $(\epsilon_{nt'})_{t'\leq t-h}$. On the other hand, $X_{nt}^{(\lfloor h/2\rfloor)}$ is a measurable function of $(\epsilon_{nt'})_{t'\geq t-\lfloor h/2\rfloor}$. Thus, we have for $h \in \mathbb{N}$

$$\left|\mathrm{Cov}\left(X_{nt}^{(\lfloor h/2\rfloor)}, X_{nt-h}\right)\right| \leq \alpha_{\epsilon_n}(\lfloor h/2\rfloor).$$

Note that α_{ϵ_n} is the mixing coefficient of the process (ϵ_{nt}). Therefore, with Assumption **B5**, we obtain

$$\sup_{n\in\mathbb{N}} \left|\mathrm{Cov}\left(X_{nt}^{(\lfloor h/2\rfloor)}, X_{nt-h}\right)\right| < K\rho^h. \tag{3.7.41}$$

Now with (3.7.40) and (3.7.41) we obtain $\sup\limits_{n\in\mathbb{N}} |\mathrm{Cov}\,(X_{nt}, X_{nt-h})| < K\rho^h$. We have obtained that $\sum_{n\geq 1} \mathrm{Var} S_n < +\infty$.

Now, using the Tchebychev's Inequality, we obtain

$$\sum_{n\geq 1} \mathbb{P}\left[|S_n - m_{n^2}| > \varepsilon\right] \leq \frac{1}{\varepsilon^2}\sum_{n\geq 1} \mathrm{Var} S_n \tag{3.7.42}$$

Thus, since the series of equation (3.7.42) is convergent we obtain the almost-sure convergence of $S_n - m_{n^2}$ to 0. We also have from Lemma 3.7.7 the almost sure convergence of m_n to $E(\inf\limits_{\tau\in V} l_t(\tau))^+$, therefore

$$S_n \to E(\inf_{\tau\in V} l_t(\tau))^+, \ a.s.$$

We now prove that M_n converges also to $E\left[(\inf\limits_{\tau\in V} l_t(\tau))^+\right]$ almost surely. Let $q_n = \lfloor\sqrt{n}\rfloor$ be the floor function of $\sqrt{n}$. Since the element of the sum M_n are positives, we have

$$\frac{1}{n} q_n^2 S_{q_n} \leq M_n \leq \frac{1}{n}(q_n+1)^2 S_{q_n+1}.$$

Using the fact that $\frac{q_n^2}{n}$ converges to 1, we obtain the $M_n \to E(\inf\limits_{\tau\in V} l_t(\tau))^+,\ a.s.$ Finally, using the same method for the negative part, we can conclude and obtain (3.7.38). □

Lemma 3.7.9. *Under the assumptions of Theorem 3.4.1, we have*

$$\lim_{n\to+\infty} \sup_{\tau\in\Gamma} \left| I_n(\tau) - \tilde{I}_n(\tau) \right| < 0, \; a.s. \tag{3.7.43}$$

Proof. Let $\underline{\tilde{\sigma}}^2_{nt}$ be the vector obtained by replacing σ^2_{nt-i} by $\tilde{\sigma}^2_{nt-i}$ and let $\underline{\tilde{c}}_{nt}$ be the vector obtained by replacing $\epsilon^2_{n0}, \dots, \epsilon^2_{n1-q}$ by some initial values. With (3.7.19), we have

$$\left|\underline{\sigma}^2_{nt} - \underline{\tilde{\sigma}}^2_{nt}\right| = \left| \sum_{k=1}^{q} B^{t-k}(\underline{c}_{nk} - \underline{\tilde{c}}_{nk}) + B^t(\underline{\sigma}^2_{n0} - \underline{\tilde{\sigma}}^2_{n0}) \right|.$$

With Assumption **B2**, we have $\sup_{\theta\in\Theta} \rho(B) < 1$, therefore

$$\sup_{\theta\in\Theta} \left|\underline{\sigma}^2_{nt} - \underline{\tilde{\sigma}}^2_{nt}\right| \le K\rho^t \left(\max(\epsilon^2_{n0}, \dots, \epsilon^2_{n1-q}) + \max(\sigma^2_{n0}, \dots, \sigma^2_{n1-p}) + 1 \right).$$

Then, since the random variables $\max(\epsilon^2_{n0}, \dots, \epsilon^2_{n1-q})$ and $\max(\sigma^2_{n0}, \dots, \sigma^2_{n1-p})$ possess moments of order s, with Lemma 3.7.5, we can conclude as we did in the proof of Theorem 3.3.1. □

Then, the proof can be done in the exact same way as in the proof of Theorem 3.3.1. Starting from Equation (3.7.3), we can use Lemma 3.7.8 to conclude.

3.7.4 Proof of the asymptotic normality in Theorem 3.4.1

We introduce a truncated version of the derivatives of σ^2_{nt}. From (3.7.19), we obtain

$$\begin{aligned}
\frac{\partial \sigma^2_{nt}}{\partial \omega}(\theta) &= \sum_{k=0}^{+\infty} B^k(1,1) \\
\frac{\partial \sigma^2_{nt}}{\partial a_i}(\theta) &= \sum_{k=0}^{+\infty} B^k(1,1)\epsilon^2_{nt-k-i}, \quad \forall i \in \{1,\dots,q\} \\
\frac{\partial \sigma^2_{nt}}{\partial b_j}(\theta) &= \sum_{k=1}^{+\infty} \left[\sum_{i=1}^{k} B^{i-1} B^{(j)} B^{k-1} \underline{c}_{nt-k}(\theta) \right](1), \quad \forall j \in \{1,\dots,p\}.
\end{aligned}$$

For $m \in \mathbb{N}$, we define

$$\left(\frac{\partial \sigma_{nt}^2}{\partial \omega}\right)^{(m)}(\theta) = \sum_{k=0}^{m} B^k(1,1) \tag{3.7.44}$$

$$\left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)}(\theta) = \sum_{k=0}^{m} B^k(1,1)\underline{z}_{nt-k-i}^{(m)}(q+1)\eta_{nt-k-i}^2, \quad \forall i \in \{1,\dots,q\} \tag{3.7.45}$$

$$\left(\frac{\partial \sigma_{nt}^2}{\partial b_j}\right)^{(m)}(\theta) = \sum_{k=1}^{m}\left[\sum_{i=1}^{k} B^{i-1}B^{(j)}B^{k-1}\underline{c}_{nt-k}^{(m)}(\theta)\right](1), \quad \forall j \in \{1,\dots,p\}\,. \tag{3.7.46}$$

where $B^{(j)}$ is a $p \times p$ matrix with 1 in position $(1,j)$ and zeros elsewhere. Then, we define

$$\phi_{nt}(\theta) = \frac{1}{\sigma_{nt}^2(\theta)}\frac{\partial \sigma_{nt}^2}{\partial \theta}(\theta),\ \phi_{nt}^{(m)}(\theta) = \frac{1}{\sigma_{nt}^{2(m)}(\theta)}\left(\frac{\partial \sigma_{nt}^2}{\partial \theta}\right)^{(m)}(\theta),$$

and for $i \in \{1,\dots,p+q+1\}$, $\phi_{nt,i}^{(m)}(\theta) = \frac{1}{\sigma_{nt}^{2(m)}(\theta)}\left(\frac{\partial \sigma_{nt}^2}{\partial \theta_i}\right)^{(m)}(\theta)$.

Lemma 3.7.10. *Under the assumptions of Theorem 3.4.1, there exists a neighborhood $V(\theta_0)$ of θ_0 such that, for any $\theta \in V(\theta_0)$ and for $(i,j,k) \in \{1,\dots,p+q+1\}$, we have*

$$\sup_{n\in\mathbb{N}} E\left|\phi_{nt,i}(\theta) - \phi_{nt,i}^{(m)}(\theta)\right| < K\rho^m \tag{3.7.47}$$

$$\sup_{n\in\mathbb{N}} E\left|\phi_{nt,i}(\theta)\phi_{nt,j}(\theta) - \phi_{nt,i}^{(m)}(\theta)\phi_{nt,j}^{(m)}(\theta)\right| < K\rho^m \tag{3.7.48}$$

$$\sup_{n\in\mathbb{N}} E\left|\phi_{nt,i}(\theta)\phi_{nt,j}(\theta)\phi_{nt,k}(\theta) - \phi_{nt,i}^{(m)}(\theta)\phi_{nt,j}^{(m)}(\theta)\phi_{nt,k}^{(m)}(\theta)\right| < K\rho^m. \tag{3.7.49}$$

And

$$\sup_{\theta\in V(\theta_0)}\sup_{n\in\mathbb{N}} E\left|\phi_{nt,i}(\theta)\right| < +\infty, \qquad \sup_{\theta\in V(\theta_0)}\sup_{n\in\mathbb{N}} E\left|\phi_{nt,i}(\theta)\phi_{nt,j}(\theta)\right| < +\infty \tag{3.7.50}$$

$$\sup_{\theta\in V(\theta_0)}\sup_{n\in\mathbb{N}} E\left|\phi_{nt,i}(\theta)\phi_{nt,j}(\theta)\phi_{nt,k}(\theta)\right| < +\infty. \tag{3.7.51}$$

Proof. I this proof, for clarity purpose, the arguments (θ) are omitted (ϕ_{nt} stands for $\phi_{nt}(\theta)$).

We have for $n \in \mathbb{N},\ t \in \mathbb{Z},\ \theta \in \Theta$ and $i \in \{1, \dots, q\}$, we have

$$\left|\frac{1}{\sigma_{nt}^2}\frac{\partial \sigma_{nt}^2}{\partial a_i} - \frac{1}{\sigma_{nt}^{2(m)}}\left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)}\right| \leq \frac{1}{\sigma_{nt}^2}\left|\frac{\partial \sigma_{nt}^2}{\partial a_i} - \left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)}\right| + \left|\left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)}\right|\left|\frac{1}{\sigma_{nt}^2} - \frac{1}{\sigma_{nt}^{2(m)}}\right|. \tag{3.7.52}$$

We begin by the first term of the previous equation, we have

$$\frac{\partial \sigma_{nt}^2}{\partial a_i} - \left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)} = \sum_{k=m+1}^{+\infty} B^k(1,1)\epsilon_{nt-k-i}^2.$$

Then, we remark that $a_i\epsilon_{nt-k-i}^2 < \underline{c}_{nt-k}(1)$ and that $\sigma_{nt}^2 > \omega + B^k(1,1)\underline{c}_{nt-k}(1)$ and obtain

$$\begin{aligned}\frac{1}{\sigma_{nt}^2}\left|\frac{\partial \sigma_{nt}^2}{\partial a_i} - \left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)}\right| &\leq \sum_{k=m+1}^{+\infty} \frac{1}{a_i}\frac{B^k(1,1)\underline{c}_{nt-k}(1)}{\omega + B^k(1,1)\underline{c}_{nt-k}(1)} \\ &\leq \sum_{k=m+1}^{+\infty} \frac{1}{a_i}\left\{\frac{B^k(1,1)\underline{c}_{nt-k}(1)}{\omega}\right\}^s,\end{aligned}$$

using the inequality $x/(1+x) \leq x^s$ for all $x \geq 0$. With Assumption **B6**, we have $a_{0i} > 0$, thus there exists a neighborhood $V(\theta_0)$ of θ_0 such that $\inf_{\theta \in V(\theta_0)} a_i > 0$. Then using Lemma 3.7.5 and the fact that the spectral radius of B is inferior to 1, we obtain

$$\sup_{n \in \mathbb{N}} E\left[\frac{1}{\sigma_{nt}^2}\left|\frac{\partial \sigma_{nt}^2}{\partial a_i} - \left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)}\right|\right] \leq K\rho^m. \tag{3.7.53}$$

Turning to the second term of Equation (3.7.52), we have by the mean value theorem applied to the function $x \mapsto x^{-1/s}$

$$\left|\frac{1}{\sigma_{nt}^2} - \frac{1}{\sigma_{nt}^{2(m)}}\right| \leq K\frac{1}{\tilde{\sigma}^{1/s+1}}\left|\sigma_{nt}^{2s} - \sigma_{nt}^{2(m)s}\right|,$$

where $\tilde{\sigma}$ is between σ_{nt}^{2s} and $\sigma_{nt}^{2(m)s}$. Since $\tilde{\sigma} \geq \sigma_{nt}^{2(m)s}$, we have

$$\left|\frac{1}{\sigma_{nt}^2} - \frac{1}{\sigma_{nt}^{2(m)}}\right| \leq K\frac{1}{\sigma_{nt}^{2(m)+2s}}\left|\sigma_{nt}^{2s} - \sigma_{nt}^{2(m)s}\right|. \tag{3.7.54}$$

Then, we have

$$a_i \left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)} = \sum_{k=0}^{m} B^k(1,1) a_i \epsilon_{nt-k-i}^2 \leq \sum_{k=0}^{m} B^k(1,1) \underline{c}_{nt-k-i}(1) = \sigma_{nt}^{2(m)},$$

and thus $\frac{1}{\sigma_{nt}^{2(m)}}\left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)} \leq K$. Now, with (3.7.54) and using Lemma 3.7.6, we obtain

$$\begin{aligned} \sup_{n\in\mathbb{N}} E\left[\left|\left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)}\right|\left|\frac{1}{\sigma_{nt}^2} - \frac{1}{\sigma_{nt}^{2(m)}}\right|\right] &\leq K \sup_{n\in\mathbb{N}} E\left[\left|\frac{1}{\sigma_{nt}^{2(m)}}\left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)}\right| \frac{1}{\sigma_{nt}^{2(m)s}}\left|\sigma_{nt}^{2s} - \sigma_{nt}^{2(m)s}\right|\right] \\ &\leq K\rho^m. \end{aligned}$$

Finally with the previous equation and (3.7.53), we obtain

$$\sup_{n\in\mathbb{N}} E\left|\frac{1}{\sigma_{nt}^2}\frac{\partial \sigma_{nt}^2}{\partial a_i} - \frac{1}{\sigma_{nt}^{2(m)}}\left(\frac{\partial \sigma_{nt}^2}{\partial a_i}\right)^{(m)}\right| < K\rho^m.$$

If we adapt the proof for the derivatives with respect to b_j and ω, we obtain (3.7.47). Now for the first part of (3.7.50), we have, for any $n \in \mathbb{N}$ with already used arguments

$$\begin{gathered} \frac{1}{\sigma_{nt}^2}\frac{\partial \sigma_{nt}^2}{\partial a_i} \leq K, \quad \frac{1}{\sigma_{nt}^2}\frac{\partial \sigma_{nt}^2}{\partial \omega} \leq K \\ E\left[\frac{1}{\sigma_{nt}^2}\frac{\partial \sigma_{nt}^2}{\partial b_j}\right] \leq \frac{1}{b_j}\sum_{k=1}^{+\infty} kE\left\{\frac{B^k(1,1)\underline{c}_{nt-k}(1)}{\omega}\right\}^s \leq \frac{K}{b_j}. \end{gathered}$$

And the first part of (3.7.50) comes easily.

Turning to (3.7.48), we have for $(i,j) \in \{1,\ldots,p+q+1\}^2$

$$\left|\phi_{nt,i}\phi_{nt,j}' - \phi_{nt,i}^{(m)}\phi_{nt,j}^{(m)\prime}\right| \leq |\phi_{nt,i}| \left|\phi_{nt,j}^{(m)} - \phi_{nt,j}^{(m)\prime}\right| + |\phi_{nt,i} - \phi_{nt,i}| \left|\phi_{nt,j}^{(m)\prime}\right|.$$

With (3.7.47) and the first part of (3.7.50), we obtain $\sup_{\theta\in V(\theta_0)}\sup_{n\in\mathbb{N}}\left|\phi_{nt,j}^{(m)}\right| < +\infty$ and (3.7.48). All the other results of the lemma can be obtained with similar arguments. $\square$

Lemma 3.7.11. *Defining*

$$\left(\frac{\partial l_{nt}}{\partial \tau}\right)^{(m)}(\tau_0) = \begin{pmatrix} \frac{1}{2}\phi_{nt}^{(m)}\left(1+\eta_{nt}\frac{\partial \log f}{\partial x}(\eta_{nt},\psi_0)\right) \\ -\frac{\partial \log f}{\partial \psi}(\eta_{nt},\psi_0) \end{pmatrix},$$

and under the assumptions of Theorem 3.4.1, we have

$$\sup_{n\in\mathbb{N}} E\left\|\frac{\partial l_{nt}}{\partial \tau}(\tau_0) - \left(\frac{\partial l_{nt}}{\partial \tau}\right)^{(m)}(\tau_0)\right\| < K\rho^m. \tag{3.7.55}$$

Proof. From (3.7.5), we have

$$\frac{\partial l_{nt}}{\partial \theta}(\tau_0) - \left(\frac{\partial l_{nt}}{\partial \theta}(\tau_0)\right)^{(m)} = \frac{1}{2}\left(1+\eta_{nt}\frac{\partial \log f}{\partial x}(\eta_{nt},\psi_0)\right)\left(\phi_{nt}-\phi_{nt}^{(m)}\right).$$

Now, since $\left(1+\eta_{nt}\frac{\partial \log f}{\partial x}(\eta_{nt},\psi_0)\right)$ only depends on η_{nt}, we can apply the dominated convergence theorem with Assumption **B3** and obtain

$$\lim_{n\to+\infty} E\left[1+\eta_{nt}\frac{\partial \log f}{\partial x}(\eta_{nt},\psi_0)\right] = E\left[1+\eta_{t}\frac{\partial \log f}{\partial x}(\eta_{t},\psi_0)\right] < +\infty.$$

The last inequality has been proved in Section 3.7.2. Since the function $x \mapsto \left(1+x\frac{\partial \log f}{\partial x}(x,\psi_0)\right)$ is bounded, it is now clear that

$$\sup_{n\in\mathbb{N}} E\left[1+\eta_{nt}\frac{\partial \log f}{\partial x}(\eta_{nt},\psi_0)\right] < +\infty.$$

Then, with Lemma 3.7.10 we obtain

$$\sup_{n\in\mathbb{N}} E\left\|\frac{\partial l_{nt}}{\partial \theta}(\tau_0) - \left(\frac{\partial l_{nt}}{\partial \theta}(\tau_0)\right)^{(m)}\right\| < K\rho^m.$$

Finally since $\left(\frac{\partial l_{nt}}{\partial \psi}(\tau_0)\right)^{(m)} = \frac{\partial l_{nt}}{\partial \psi}(\tau_0)$ we obtain (3.7.55). □

Lemma 3.7.12. *Under the assumptions of Theorem 3.4.1, we have*

$$\lim_{n\to+\infty} E\frac{\partial l_{nt}}{\partial\tau}(\tau_0) = E\frac{\partial l_t}{\partial\tau}(\tau_0) \quad and \quad \lim_{n\to+\infty} E\left[\frac{\partial l_{nt}}{\partial\tau}(\tau_0)\frac{\partial l_{nt}}{\partial\tau'}(\tau_0)\right] = E\left[\frac{\partial l_t}{\partial\tau}(\tau_0)\frac{\partial l_t}{\partial\tau'}(\tau_0)\right]. \tag{3.7.56}$$

Proof. To obtain the result, we first prove that, for any $m \in \mathbb{N}$, $E\phi_{nt}^{(m)}(\theta_0) \to E\phi_t^{(m)}(\theta_0)$ when $n \to +\infty$. From (3.7.44), we know that $\left(\frac{\partial\sigma_{nt}^2}{\partial\omega}\right)^{(m)}(\theta_0)$ does not depend on n. Since $\sigma_{nt}^2 \geq \omega_0 > 0$, we can apply the dominated convergence theorem and obtain

$$E\phi_{nt,1}^{(m)}(\theta_0) \to E\phi_{t,1}^{(m)}(\theta_0), \text{ when } n \to +\infty.$$

Now, using the same method as in the proof of Lemma 3.7.10, we obtain for $i \in \{1, \cdots, q\}$

$$\phi_{nt,1+i}^{(m)} \leq \sum_{k=0}^{m} \frac{1}{a_{0,i}} \left\{\frac{B^k(1,1)\underline{c}_{nt-k}^{(m)}(1)}{\omega_0}\right\}^s,$$

where s can be chosen as small as wanted. With the same arguments as in the proof of Lemma 3.7.7 and the previous equation, we obtain that $\phi_{nt,1+i}^{(m)}$ is a function of $\{\eta_{nt-k}; 1 \leq k \leq 2m+q\}$ and is such that for any $s > 0$ there exist $K, M > 0$ such that $\phi_{nt,1+i}^{(m)} \leq K\prod_{i=1}^{2m+q}\max\left(M, \eta_{nt-i}^s\right)$. Then, with the dominated convergence theorem, we obtain

$$E\phi_{nt,1+i}^{(m)}(\theta_0) \to E\phi_{t,1+i}^{(m)}(\theta_0), \text{ when } n \to +\infty.$$

Doing exactly the same for $\phi_{nt,1+q+j}^{(m)}$ (as in Lemma 3.7.10) with $j \in \{1, \ldots, p\}$, we obtain

$$E\phi_{nt}^{(m)}(\theta_0) \to E\phi_t^{(m)}(\theta_0), \text{ when } n \to +\infty.$$

Since $1 + \eta_{nt}\frac{\partial \log f}{\partial x}(\eta_{nt}, \psi_0)$ only depends on η_{nt}, we easily obtain

$$E\left[1 + \eta_{nt}\frac{\partial \log f}{\partial x}(\eta_{nt}, \psi_0)\right] \underset{n\to+\infty}{\longrightarrow} 0.$$

Therefore we have

$$\lim_n E\left(\frac{\partial l_{nt}}{\partial\theta}\right)^{(m)}(\theta_0) = E\left(\frac{\partial l_t}{\partial\theta}\right)^{(m)}(\theta_0).$$

Now, with the asymptotic expansions (3.2.8)-(3.2.10) and with previously used arguments, we also obtain the convergence for the derivatives with respect to ψ. Finally, inverting the double limit $\lim_n \lim_m E\left(\frac{\partial l_{nt}}{\partial \tau}\right)^{(m)}(\theta_0) = \lim_m \lim_n E\left(\frac{\partial l_{nt}}{\partial \tau}\right)^{(m)}(\theta_0)$, we obtain the first part of (3.7.56). It is clear that the second part of (3.7.56) can be obtained with very similar arguments. $\square$

Lemma 3.7.13. *Under the assumptions of Theorem 3.4.1, we have*

$$\sqrt{n}\frac{\partial I_n}{\partial \tau}(\tau_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, J), \tag{3.7.57}$$

with $J = E\left[\frac{\partial l_t}{\partial \tau}(\tau_0)\frac{\partial l_t}{\partial \tau'}(\tau_0)\right]$.

Proof. For $\lambda \in \mathbb{R}^{p+q+4}$, we define $Y_{nt} = \lambda' \frac{\partial l_{nt}}{\partial \tau}(\tau_0)$, $Y_t = \lambda' \frac{\partial l_t}{\partial \tau}(\tau_0)$ and $z_{nt} = \frac{1}{\sqrt{n}}\left(Y_{nt} - EY_{nt}\right)$. We will apply the central limit theorem of Lindeberg to the array (z_{nt}) to prove this lemma. We obviously have $Ez_{nt} = 0$ and $E\left[z_{nt}^2\right] = \frac{1}{n}\left(E\left[Y_{nt}^2\right] - E\left[Y_{nt}\right]^2\right)$. Now, with Lemma 3.7.12, we have, when n tends to infinity

$$E\left[Y_{nt}^2\right] \to E\left[Y_t^2\right] < +\infty, \text{ and } E\left[Y_{nt}\right]^2 \to E\left[Y_t\right]^2 = 0,$$

using the results of the proof of Lemma 3.7.1. The Lindeberg condition remains to be proven. We have for any $\varepsilon > 0$,

$$\sum_{t=1}^{n} E\left[z_{nt}^2 \mathbb{1}_{|z_{nt}|>\varepsilon}\right] = \sum_{t=1}^{n} \frac{1}{n}\int_{|Y_{nt}-EY_{nt}|>\varepsilon\sqrt{n}} |Y_{nt} - EY_{nt}|^2 dP,$$

and $P\left[|Y_{nt} - EY_{nt}| > \varepsilon\sqrt{n}\right] \to 0$, when $n \to +\infty$. Besides, we have $\sup_{n\in\mathbb{N}} E\left|Y_{nt} - EY_{nt}\right|^2 < +\infty$. We can conclude and obtain the Lindeberg condition

$$\sum_{t=1}^{n} E\left[z_{nt}^2 \mathbb{1}_{|z_{nt}|>\varepsilon}\right] \to 0, \text{ when } n \to +\infty.$$

It remains to apply the Lindeberg central limit theorem and the Wold-Cramer theorem and we obtain (3.7.57). $\square$

Lemma 3.7.14. *Under the assumptions of Theorem 3.4.1, we have*

$$\frac{\partial^2}{\partial \tau \partial \tau'} I_n(\tau_0) \to -J, \quad a.s. \tag{3.7.58}$$

Proof. Adapting (3.7.10) and (3.7.11), we obtain for $(i,j) \in \{1,\dots,p+q+1\}$

$$\begin{aligned}\frac{\partial^2 l_{nt}}{\partial\theta_i\partial\theta_j} &= \frac{1}{2}\left(\frac{1}{\sigma_{nt}^2}\frac{\partial^2\sigma_{nt}^2}{\partial\theta_i\theta_j} - \phi_{nt,i}\phi_{nt,j}\right)\left(1+\eta_{nt}\frac{\partial\log f}{\partial x}\right) \\ &\quad -\frac{1}{4}\phi_{nt,i}\phi_{nt,j}\eta_{nt}\left(\frac{\partial\log f}{\partial x}+\eta_{nt}\frac{\partial^2\log f}{\partial x^2}\right) \\ \frac{\partial^2 l_{nt}}{\partial\theta_i\partial\psi} &= \frac{1}{2}\phi_{nt,i}\eta_t\frac{\partial^2\log f}{\partial x\partial\psi}\end{aligned}$$

Using (7.46) and (7.47) in Francq and Zakoian (2010) and the same reasoning as in Lemma 3.7.7 (defining a truncated version of $\partial^2\sigma_{nt}^2/\partial\tau\partial\tau'$), we obtain for $(i,j) \in \{1,\dots,p+q+1\}^2$

$$E\left[\frac{\partial^2 l_{nt}}{\partial\tau_i\partial\tau_j}(\tau_0)\right] \underset{n\to+\infty}{\longrightarrow} E\left[\frac{\partial^2 l_t}{\partial\tau_i\partial\tau_j}(\tau_0)\right].$$

Then, as in Lemma 3.7.8, we obtain the result. □

Lemma 3.7.15. *Under the assumptions of Theorem 3.4.1, there exists a neighborhood $V(\tau_0)$ of τ_0 such that for $(i,j,k) \in \{1,\dots,p+q+1\}^3$*

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\tau\in V(\tau_0)}\frac{\partial^3 l_{nt}}{\partial\tau_i\partial\tau_j\partial\tau_k}(\tau) \to E\left[\sup_{\tau\in V(\tau_0)}\frac{\partial^3 l_t}{\partial\tau_i\partial\tau_j\partial\tau_k}(\tau)\right], \quad a.s. \tag{3.7.59}$$

Proof. Using the results of Lemma 3.7.10 and of Lemma 3.7.2 the proof is straightforward. □

Lemma 3.7.16. *Under the assumptions of Theorem 3.4.1, we have when $n \to +\infty$*

$$\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{\partial l_{nt}}{\partial\tau}(\tau_0) - \frac{\partial\tilde{l}_{nt}}{\partial\tau}(\tau_0)\right\}\right\| \to 0 \tag{3.7.60}$$

$$\sup_{\tau\in\Gamma}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial^2 l_{nt}}{\partial\tau\partial\tau'}(\tau) - \frac{\partial^2\tilde{l}_nt}{\partial\tau\tau'}(\tau)\right\}\right\| \to 0 \tag{3.7.61}$$

Proof. This lemma can be easily proven using the same arguments as in Lemma 3.7.3 and Lemma 3.7.9. □

Proof of Theorem 3.4.1. Using Lemmas 3.7.10-3.7.16 and the same method as in the proof of the asymptotic normality in the ML case, we obtain the result. □

3.7.5 Proof of Theorem 3.4.2

With a Taylor expansion, we obtain for $(i,j) \in \{1,\dots,p+q+4\}^2$

$$\frac{\partial^2 l_{nt}}{\partial\tau_i\partial\tau_j}(\tau_n) = \frac{\partial^2 l_{nt}}{\partial\tau_i\partial\tau_j}(\tau_0) + \frac{\partial^3 l_{nt}}{\partial\tau'\partial\tau_i\partial\tau_j}(\tilde{\tau})(\tau_n - \tau_0).$$

Using Lemma 3.7.2, Lemma 3.7.15, the equivalent of Lemma 3.7.12 for the second order derivatives and the consistency of the estimator τ_n, we obtain

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 l_{nt}}{\partial\tau_i\partial\tau_j}(\tau_n) \underset{n\to+\infty}{\longrightarrow} E\left[\frac{\partial^2 l_t}{\partial\tau_i\partial\tau_j}(\tau_0)\right].$$

Bibliography

S.K. Basu and M. Maejima. A local limit theorem for attractions under a stable law. 87(01): 179–187, 1980.

H. Bergström. On some expansions of stable distributions. *Arkiv for Matematik*, 2(3):7, 1952.

I. Berkes and L. Horváth. The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics*, 32(2):633–655, 2004.

I. Berkes, L. Horvath, and P. Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, pages 201–227, 2003.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

T. Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, pages 542–547, 1987.

T. Bollerslev. Glossary to ARCH (GARCH). *CREATES Research Paper*, 49, 2008.

P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52(1-2):115–127, 1992. ISSN 0304-4076.

F. Boussama. Ergodicité, mélange et estimation dans les modèles GARCH. Doctoral thesis, Université Paris 7. 1998.

J.M. Chambers, C.L. Mallows, and BW Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, pages 340–344, 1976.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events for insurance and finance*, volume 33. Springer Verlag, 1997.

R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, pages 987–1007, 1982.

E.F. Fama. The behavior of stock-market prices. *The journal of Business*, 38(1):34–105, 1965.

C. Francq and J.M. Zakoïan. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–638, 2004.

C. Francq and J.M. Zakoian. *GARCH models: Structure, statistical inference and financial applications.* Wiley, 2010.

C. Francq, G. Lepage, and J.M. Zakoïan. Two-stage non Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics*, 165(2): 246–257, 2011.

B.V. Gnedenko, A.N. Kolmogorov, K.L. Chung, and J.L. Doob. *Limit distributions for sums of independent random variables.* Addison-Wesley Reading, MA, 1968.

S.M. Liu and B.W. Brorsen. Maximum likelihood estimation of a GARCH-stable model. *Journal of Applied Econometrics*, 10(3):273–285, 1995.

B. Mandelbrot. The variation of certain speculative prices. *The journal of business*, 36(4):394–419, 1963. ISSN 0021-9398.

M. Matsui and A. Takemura. Some improvements in numerical evaluation of symmetric stable density and its derivatives. *Communications in Statistics-Theory and Methods*, 35(1):149–172, 2006.

S. Mittnik and M.S. Paolella. Prediction of financial downside-risk with heavy-tailed conditional distributions. 2003.

J.P. Nolan. Numerical calculation of stable densities and distribution functions. *Stochastic Models*, 13(4):759–774, 1997.

G. Samorodnitsky and M.S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance.* Chapman & Hall/CRC, 1994. ISBN 0412051710.

N. Shephard. Statistical aspects of ARCH and stochastic volatility. *Time series models in econometrics, finance and other fields*, pages 1–67, 1996.

V.M. Zolotarev. *One-dimensional stable distributions.* American Mathematical Society, 1986. ISBN 0821845195.

Chapter 4

Inference for a Misspecified MS-GARCH process

4.1 Introduction

Since the seminal papers of Engle (1982) and Bollerslev (1986), the Generalized AutoRegressive Conditionally Heteroskedastic (GARCH) models have been widely used to explain and forecast the volatility of financial time series. The reason is that this model explains some of the stylized facts that can be found in financial series (the volatility clustering, the leptokurticity, ...).

This takes into account the stylized fact that most financial assets exhibit a volatility clustering, i.e. large movements tend to be followed by other large movements. The GARCH(1,1) model can be written as follows

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \end{cases} \tag{4.1.1}$$

where (η_t) is a white noise with unit variance and $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$. This model is widely used among practitioners and researchers and can be seen as the default model used to explain financial time series. Since this model is almost systematically used, one can argue that sometimes it will be applied to other Data Generating Processes (DGP) and that the model will be misspecified. Therefore, it is useful to know how the usual estimators behave when a GARCH model is inferred on a different DGP.

On large periods of time, the GARCH model tends to be insufficient to explain the dynamics of the volatility. Indeed, some researchers have found strong evidences for the occurrence of structural breaks, see for exemple Andreou and Ghysels (2002). In this case, the structural breaks may cause the unconditional variance of the process to vary in time: this is for instance the case if the parameters change over time. Mikosch and Stărică (2004) and Hillebrand (2005) have studied GARCH models with structural breaks. Another way to allow for apparent breaks, while remaining in the stationary framework is to use Markov Switching (MS) models. Hamilton and Susmel (1994) have introduced an ARCH model where the dynamic of the coefficients is governed by a Markov chain. This model and its generalization to MS-GARCH model have received considerable attention, see Cai (1994), Dueker (1997), Gray (1996), Francq and Zakoïan (2005) and Bauwens et al. (2010). The MS-GARCH model is more flexible than the classical GARCH model, the regime changes can be interpreted as different economic phases. The principal issue with the MS-GARCH model is that its estimation is extremely difficult. The marginal likelihood of a MS-ARCH model can be computed, but the algorithm used by Fruhwirth-Schnatter and Kaufmann (2002) on this model cannot be extended to the MS-GARCH model because of the path dependence problem. Some papers studied the inference of MS-GARCH models with numerical methods such as Gibbs sampling, see for instance Bauwens et al. (2010) or Henneke et al. (2011). One of the issues of these methods is that they do not yield an asymptotic law for the estimators.

In this paper, we study the asymptotic behavior of the estimates of a GARCH(1,1) model when the DGP is a MS-GARCH(p,q). We consider the most widely used estimator for a GARCH(1,1) model that is the Gaussian Quasi Maximum Likelihood Estimator (QMLE). In the well specified case (in the absence of regime switches), this estimator is consistent and asymptotically normal under mild assumptions. The main question is whether this estimator converges toward a pseudo true value and the information that this false model can provide.

The estimation of misspecified models has been studied by White (1982). Most of the papers related to this topic studied a misspecification on the distribution of the innovation process, see for example Gourieroux et al. (1984) or Domowitz and White (1982). Other papers have been devoted to a misspecification in the model, see for example Tanaka and Maekawa (1984), Mevel and Finesso (2004), Ogata (1980). The estimation of misspecified GARCH models on other DGPs has been studied by Jensen and Lange (2010), the found that, when the DGP is a continuous time stochastic volatility model and when the sampling frequency tends to infinity, then the QMLE of a GARCH(1,1) model converges to $(0, 0, 1)$. In this paper, both the DGP and the estimated model are discrete and we find the convergence of the misspecified estimates to a pseudo true value which depends on the DGP.

The main objective of this paper is to present to theoretical results of the inference of a GARCH(1,1) model on a MS-GARCH(p,q) model. In a first section, we will present some properties of the MS-GARCH model and in particular, we will give a new condition of stationarity. Then we will prove the consistency and the asymptotic normality of the QMLE in the case where the DGP is stationary. In the line of Francq and Zakoïan (2012), we will also in a third part study the asymptotic behavior of this estimator in the non-stationary case and prove the convergence of the estimators of parameters α and β. We will then present a simulation study to illustrate the convergence toward pseudo true values and to link this framework with the work of Mikosch and Stărică (2004) and Hillebrand (2005). Most of the proofs are deferred to a last section.

4.2 The MS-GARCH model

The existence of strictly stationary, and second order stationary solutions to MS-GARCH models was established by Francq et al. (2001), see also Liu (2006). In this section, we give complementary results on the probability structure of the MS-GARCH model. We consider a process $(\epsilon_t)_{t\in\mathbb{Z}}$ satisfying an equation of the form

$$\begin{cases} \epsilon_t = \sqrt{h_t} u_t \\ h_t = w(\Delta_t) + \sum_{i=1}^{q} a_i(\Delta_t)\epsilon_{t-i}^2 + \sum_{j=1}^{p} b_j(\Delta_t) h_{t-j}, \end{cases} \tag{4.2.1}$$

where (ϵ_t) is the observed process ($\epsilon_t \in \mathbb{R}$), (u_t) is the error process which is independent and identically distributed and such that $Eu_t = 0$ and $Eu_t^2 = 1$ (we denote (u_t) iid$(0,1)$). The process (Δ_t) is a finite state space ergodic Markov chain, which is independent of the sequence (u_t). Let $\Delta_t \in \{1, 2, \cdots, d\}$. The coefficients are such that

$$\forall k \in \{1, \cdots, d\}, \ \forall (i,j) \in \{1, \cdots, q\} \times \{1, \cdots, p\}, \ w(k) > 0, \ a_i(k) \geq 0, \ b_j(k) \geq 0.$$

The stationary probabilities of (Δ_t) are denoted by $\pi(k) = \mathbb{P}[\Delta_1 = k]$. The transition probabilities are denoted by $p(k,l) = \mathbb{P}[\Delta_t = l | \Delta_{t-1} = k]$ for $k, l \in \{1, \cdots, d\}$. For a function $f : \{1, \cdots, d\} \mapsto$

$\mathcal{M}_{n\times n'}(\mathbb{R})$, where $\mathcal{M}_{n\times n'}(\mathbb{R})$ is the set of real matrices of size $n \times n'$, we define

$$\mathbb{P}_f = \begin{pmatrix} p(1,1)f(1) & \cdots & p(d,1)f(1) \\ \vdots & & \vdots \\ p(d,1)f(d) & \cdots & p(d,d)f(d) \end{pmatrix} \quad \text{and} \quad \Pi_f = \begin{pmatrix} \pi(1)f(1) \\ \vdots \\ \pi(d)f(d) \end{pmatrix}. \tag{4.2.2}$$

We also define the backward transition probabilities,

$$\forall (i,j) \in \{1,\cdots,d\}^2, \quad q(i,j) = \mathbb{P}\left[\Delta_{t-1} = j |\ \Delta_t = i\right],$$

and the matrix $\mathbb{Q}_f$ is defined as the matrix $\mathbb{P}_f$, obtained by replacing $(p(i,j))_{i,j}$ by $(q(i,j))_{i,j}$.

Let $r = \max(p,q)$ and if $p > q$, define $b_i = 0$ for $i \in \{q+1,\ldots,p\}$ and $a_j = 0$ for $j \in \{p+1,\ldots,q\}$ if $q > p$. From (4.2.1), we write a vectorial expression of the model. Letting

$$A_t = \begin{pmatrix} a_1(\Delta_t)u_t^2 + b_1(\Delta_t) & & \cdots & & a_r(\Delta_t)u_{t-r}^2 + b_r(\Delta_t) \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & & \cdots & 1 & 0 \end{pmatrix}, \tag{4.2.3}$$

and

$$\underline{z}_t = \begin{pmatrix} h_t \\ \vdots \\ h_{t-p+1} \end{pmatrix}, \ \underline{b}_t = \begin{pmatrix} w(\Delta_t) \\ \vdots \\ 0 \end{pmatrix} \text{ and } B_t = B(\Delta_t) = \begin{pmatrix} b_1(\Delta_t) & & \cdots & & b_q(\Delta_t) \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & & 1 & 0 \end{pmatrix}. \tag{4.2.4}$$

Model (4.2.1) can be equivalently written as

$$\underline{z}_t = \underline{b}_t + A_t \underline{z}_{t-1}. \tag{4.2.5}$$

For $k \in \{1,\cdots,d\}$ and $s > 0$, we define $\tilde{A}^s(k) = E\left[A_t^{(s)} | \Delta_t = k\right]$, where for a generic matrix M, the matrix $M^{(s)} = \left(M(i,j)^s\right)_{(i,j)}$. For any square matrix M, let $\rho(M)$ denote the spectral radius of M. We state the following assumption,

A0 There exists $s \in (0,1)$ such that $\rho\left(\mathbb{P}_{\tilde{A}^s}\right) < 1$.

Lemma 4.2.1. *Under Assumption **A0**, Model* (4.2.1) *admits a unique non anticipative strictly* *stationary solution* (ϵ_t) *and for any* $r > 0$ *such that* $r < s$*, we have*

$$E|\epsilon_t|^{2r} < +\infty, \quad E\,|h_t|^r < +\infty. \tag{4.2.6}$$

In the following, our stationarity assumption is **A0**. This assumption allows us to obtain the existence of a moment or order s for the process $(\epsilon_t)_t$. Define γ as the top Lyapunov exponent associated to the sequence (A_t),

$$\gamma = \lim_{t \to +\infty} \frac{1}{t} \log \|A_t \cdots A_1\|, \quad \text{a.s.} \tag{4.2.7}$$

The classical stationarity assumption is $\gamma < 0$ which, by the Cauchy criterion ensures, that the series $\sum_{k=1}^{+\infty} A_t \cdots A_{t-k+1} \underline{b}_{t-k}$ converges and is finite.

Remark 4.2.1. In the GARCH$(1,1)$ case, **A0** implies the usual stationary assumption $\gamma < 0$. Moreover, if this assumption is not satisfied, we have

$$\gamma > 0 \Rightarrow h_t \to +\infty, \quad \text{a.s.}$$

Indeed, if **A0** is true and if $\gamma > 0$, we obtain by the Cauchy criterion

$$\sum_{k=1}^{n} A_t \cdots A_{t-k+1} \underline{b}_{t-k} \underset{n \to +\infty}{\longrightarrow} +\infty, \ a.s.$$

which yields $h_t \to +\infty$, almost surely. That is in contradiction with Lemma 4.2.1.

The case $\gamma = 0$ is more complicated. In the GARCH case, Klüppelberg et al. (2004) proved that $\gamma = 0$ implies that $h_t \to +\infty$ in probability. Their proof can be adapted to the MS-GARCH model. The convergence in probability of h_t toward $+\infty$ is in contradiction with the existence of a finite moment of order $2s$ for the process $(\epsilon_t)_t$.

4.3 Estimation of a misspecified stationary GARCH model

In the last section, we introduced the MS-GARCH model. In this model, the values of the parameters of the GARCH specification can change and depend on an unobserved Markov chain. Since this Markov chain is not observed, it is likely that the model will be misspecified and that a classic GARCH model will be inferred. In most cases, the practitioners use a GARCH(1,1) model. In this section, we study the behavior of QMLE of a GARCH(1,1) model, when the DGP is a MS-GARCH model.

The parameter of interest of Model (4.1.1) is $\theta = (\omega, \alpha, \beta)'$, which is included in a parameter space $\Theta \subset (0, +\infty) \times [0, +\infty)^2$. For any $\theta \in \Theta$, we define

$$\forall t,\ \sigma_t^2(\theta) = \sum_{k=0}^{+\infty} \beta^k (\omega + \alpha \epsilon_{t-k-1}^2). \tag{4.3.1}$$

For $\beta < 1$ and if $E|\epsilon_t|^{2s} < +\infty$, then $\sigma_t^2(\theta) \in \mathbb{R}_+$ almost surely. We define the QMLE and the criterion to be minimized

$$\tilde{I}_n(\theta) = \sum_{t=1}^{n} \tilde{l}_t(\theta), \quad \tilde{l}_t(\theta) = \log \tilde{\sigma}_t^2(\theta) + \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\theta)},$$

where the $\tilde{\sigma}_t(\theta)$ are recursively defined using some initial values. The quantities $I_n(\theta)$ and $l_t(\theta)$ are the counterparts of $\tilde{I}_n(\theta)$ and $\tilde{l}(\theta)$ in which $(\tilde{\sigma}_t(\theta))_t$ are replaced by $(\sigma_t(\theta))_t$. The Gaussian QMLE is defined as $\theta_n = \underset{\theta\in\Theta}{\operatorname{argmin}}\ \tilde{I}_n(\theta)$. In the following, for brevity purpose, we write σ_t for $\sigma_t(\theta)$.

We will prove that under appropriate assumptions, if $El_t(\theta)$ has a unique minimizer $\theta_0 \in \Theta$, then θ_n converges almost surely toward θ_0. We need the following assumptions.

A1 Θ is a compact and for all $\theta \in \Theta,\ \beta < 1$.

A2 $\rho(\mathbb{Q}_B) < 1$, where B is defined in (4.2.4).

A3 $\Theta^* \neq \emptyset$, where $\Theta^* = \{\theta^* = (\omega^*, \alpha^*, \beta^*) \in \Theta,\ \rho(\mathbb{Q}_B) < \beta^* < 1\}$.

A4 There exists a unique $\theta_0 = (w_0, \alpha_0, \beta_0)' \in \Theta$ such that $\theta_0 = \underset{\theta\in\Theta}{\operatorname{argmin}}\ El_t(\theta)$ and $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ.

Remark 4.3.1. In some cases, assuming that $\rho(\mathbb{Q}_B) < 1$ is equivalent to assuming that $\rho(\mathbb{P}_B) < 1$. We can prove that the spectrum of $\mathbb{P}_B$ and $\mathbb{Q}_B$ are identical in the case $d = 2$, see Section 4.6.

Proposition 4.3.1. *Under Assumptions **A0-A4**, for any $\theta^* \in \Theta^*$, we have*

$$E|l_t(\theta^*)| < +\infty. \tag{4.3.2}$$

Moreover, we have

$$E\,|l_t(\theta_0)| < +\infty. \tag{4.3.3}$$

With the forementioned assumptions, we are now able to obtain the convergence of the estimator θ_n toward the pseudo true value θ_0.

Theorem 4.3.1. *Under Assumptions **A0-A4**, we obtain the strong consistency of θ_n toward the* *pseudo true value θ_0, that is*

$$\theta_n \underset{n\to+\infty}{\longrightarrow} \theta_0, \ a.s. \tag{4.3.4}$$

Remark 4.3.2. The existence part of Assumption **A4** can be easily obtained if we assume that the function $\theta \mapsto El_t(\theta)$ is continuous, then using the compactness of the parameter space and Proposition 4.3.1, we can conclude.
This assumption is also easily obtained if $p = q = 1$. In this case, we can use Lemma 4.6.9 (used in the proofs of the results of Section 4.4) and the result is straightforward.

We need some additional assumptions to obtain the asymptotic normality of the misspecified pseudo maximum likelihood estimator.

A5 There exist $\eta_2 > 0$ such that $E\left|l_t(\theta_0)\right|^{1+\eta_2} < +\infty$.

A6 There exists $\eta_3 > 0$ such that $E|u_t|^{4+\eta_3} < +\infty$.

A7 The matrix $A(\theta_0)$ is invertible, where $A(\theta_0) = E\left[\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right]$.

Theorem 4.3.2. *Under Assumptions **A0-A7**, we have*

$$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1}\right), \tag{4.3.5}$$

where $B(\theta_0)$ is a well defined positive definite matrix given by

$$B(\theta_0) = E\left[\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}\right].$$

Remark 4.3.3. Assumption **A7** can be formulated as in White (1982) by assuming that θ_0 is a regular point of $A(\theta)$. A regular point of a matrix $A(\theta)$ is defined as a value of θ such that $A(\theta)$ has constant rank in some open neighborhood of θ. In this case, using a second order Taylor expansion, we can prove that for any neighborhood V of θ_0, there exists $\theta^* \in V$ such that $A(\theta^*)$ has full rank and therefore, with this assumption, $A(\theta_0)$ is invertible.

4.4 Estimation of a misspecified non stationary GARCH model

We are now interested in the case where Assumption **A0** is not satisfied and the true model is non stationary. In the well specified non stationary GARCH model, Francq and Zakoïan (2012) showed that the parameters α_0 and β_0 can be consistently estimated. In this section, we will show that under some assumptions we can demonstrate the same property in the non stationary case.

4.4.1 The MS-ARCH(1) case

We treat a simple case. The DGP is a MS-ARCH(1) model and we estimate a classic ARCH(1) model. The DGP can be written

$$\begin{cases} \epsilon_t = \sqrt{h_t} u_t, \quad (u_t) \text{ iid}(0,1) \\ h_t = w(\Delta_t) + a(\Delta_t)\epsilon_{t-1}^2. \end{cases} \tag{4.4.1}$$

For this model, the top Lyapunov exponent γ reduces to

$$\gamma = E\left[\log a(\Delta_t) u_t^2\right] = \sum_{k=1}^{d} \pi(k) \log a(k) + E \log u_t^2. \tag{4.4.2}$$

Let $\alpha_0 = E\left[a(\Delta_t)\right]$ and let $\Theta \subset (0, +\infty)^2$ be a compact parameter space such that $\alpha_0 \in \overset{\circ}{\Theta}$. Using a QML estimator again, we define the criterion to be minimized, for $\theta = (\omega, \alpha)'$

$$I_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left\{ \log \frac{\sigma_t^2}{h_t} + u_t^2 \left(\frac{h_t}{\sigma_t^2} - 1 \right) \right\}. \tag{4.4.3}$$

We define the QML estimator of the misspecified model $\theta_n = (w_n, a_n)' = \underset{\theta \in \Theta}{\operatorname{argmin}}\ I_n(\theta)$.

Theorem 4.4.1. *If $\gamma > 0$, we have*

$$\alpha_n = \frac{1}{n} \sum_{t=1}^{n} u_t^2 a(\Delta_t) + o(1), \quad a.s.\ when\ n \to +\infty. \tag{4.4.4}$$

In consequence, we have

$$\alpha_n \underset{n \to +\infty}{\longrightarrow} \alpha_0 = E\left[\alpha(\Delta_t)\right],\ a.s. \tag{4.4.5}$$

Moreover if $Eu_t^4 < +\infty$, we have

$$\sqrt{n}\left(\alpha_n - \alpha_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, E\left[a(\Delta_t)^2\right]\left(Eu_t^4 - 1\right)\right). \tag{4.4.6}$$

We emphasize the fact that in the MS-ARCH(1), we prove that the pseudo true value α_0 is unique and we explicit α_0 as a function of the parameters $a(k),\ k \in \{1, \cdots, d\}$ and of the stationary probabilities of the Markov chain $(\Delta_t)_t$.

4.4.2 The GARCH case

We are now interested in the case where the DGP is a non stationary MS-GARCH(1,1) process and we estimate a GARCH(1,1) model. The DGP can be written

$$\begin{cases} \epsilon_t = \sqrt{h_t} u_t, \quad (u_t)_t \text{ iid}(0,1) \\ h_t = w(\Delta_t) + a(\Delta_t)\epsilon_{t-1}^2 + b(\Delta_t)h_{t-1}, \ \forall t. \end{cases} \tag{4.4.7}$$

We have for this model $\gamma = E[\log a_t]$, where $a_t = a(\Delta_{t+1})u_t^2 + b(\Delta_{t+1})$. The Markov chain (Δ_t) and the process of the innovations (u_t) are as in the last sections supposed to be independent. We define $\underline{w} = \min\{w(i),\ 1 \leq i \leq d\}$ and $\overline{w} = \max\{w(i),\ 1 \leq i \leq d\}$. We define $\underline{a}$, $\overline{a}$, $\underline{b}$ and $\overline{b}$ in the same way. We give, for this model a modified version of the result given in Remark 4.2.1.

Proposition 4.4.1. *When $\gamma > 0$, $h_t \to +\infty$ at an exponential rate. We have, for any $\rho > e^{-\gamma}$, when $t \to +\infty$*

$$\rho^t h_t \to +\infty, \ \ a.s.$$

Moreover, if $E|\log u_1^2| < +\infty$, we have

$$\rho^t \epsilon_t \to +\infty, \ \ a.s.$$

This result is very similar to Proposition A.1 in Francq and Zakoïan (2012). They prove it for a GARCH model but the proof can be easily adapted to the MS-GARCH(1,1) case.

We proceed as if the true model was a GARCH(1,1) model and estimate the parameter $\theta = (\omega, \alpha, \beta)' \in \Theta$ of this model by minimizing the criterion $I_n(\theta)$ defined by

$$I_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left\{u_t^2\left(\frac{h_t}{\sigma_t^2(\theta)} - 1\right) + \log\frac{\sigma_t^2(\theta)}{h_t}\right\}, \tag{4.4.8}$$

with $\sigma_t^2(\theta) = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2(\theta)$ with initial values for ϵ_0^2 and $\sigma_0^2(\theta)$. $\Theta \subset (0, +\infty)^3$ is the parameter space. The misspecified estimator $\theta_n = (\omega_n, \alpha_n, \beta_n)'$ of the GARCH(1,1) model is defined by $\theta_n = \underset{\theta\in\Theta}{\operatorname{argmin}}\ I_n(\theta)$. We define

$$v_t(\alpha, \beta) = \sum_{j=1}^{+\infty}\alpha\beta^{j-1}u_{t-j}^2\prod_{k=1}^{j-1}\frac{1}{a_{t-k}}. \tag{4.4.9}$$

We prove in Lemma 4.6.8 that $v_t(\alpha, \beta)$ can be considered as a stationary analog of $\frac{\sigma_t^2(\theta)}{h_t}$. The criterion $I_n(\theta)$ can also be written $I_n(\theta) = O_n(\alpha, \beta) + R_n(\theta)$, where

$$O_n(\alpha, \beta) = \frac{1}{n} \sum_{t=1}^{n} o_t(\alpha, \beta) = \frac{1}{n} \sum_{t=1}^{n} \left\{ u_t^2 \left(\frac{1}{v_t(\alpha, \beta)} - 1 \right) + \log v_t(\alpha, \beta) \right\}, \tag{4.4.10}$$

and

$$R_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} r_t(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left\{ u_t^2 \left(\frac{h_t}{\sigma_t^2} - \frac{1}{v_t(\alpha, \beta)} \right) + \log \frac{\sigma_t^2}{h_t v_t(\alpha, \beta)} \right\}. \tag{4.4.11}$$

We assume

B0 $\gamma > 0$.

B1 There exists a unique minimizer (α_0, β_0) such that

$$(\alpha_0, \beta_0) = \underset{(\alpha,\beta)}{\operatorname{argmin}}\ E \left[u_t^2 \left(\frac{1}{v_t(\alpha, \beta)} - 1 \right) + \log v_t(\alpha, \beta) \right].$$

B2 The parameter space Θ is a compact set.

B3 The innovation process (u_t) is such that $\mathbb{P}[u_t = 0] = 0$.

Theorem 4.4.2. *Under Assumptions **B0-B3**, we have*

$$\alpha_n \to \alpha_0, \ \textit{and } \beta_n \to \beta_0, \ \textit{a.s. when } n \to +\infty. \tag{4.4.12}$$

Remark 4.4.1. Assumption **B1** can be formulated differently. We define, for any $\beta > 0$

$$F_t(\beta) = \frac{v_t(\alpha, \beta}{\alpha} = \sum_{j=1}^{+\infty} \beta^{j-1} u_{t-j}^2 \prod_{k=1}^{j-1} \frac{1}{a_{t-k}}.$$

We state the following alternate assumption.

B1' There exists a unique minimize $\beta_0 < e^{\gamma}$ such that

$$\beta_0 = \underset{\beta}{\operatorname{argmin}} \left\{ \log E \left[\frac{1}{F_t(\beta)} \right] - E \left[\log \frac{1}{F_t(\beta)} \right] \right\}.$$

The proof that Assumption **B1'** can be a substitute for Assumption **B1** is given in Section 4.6.

Under this assumption, the pseudo true value α_0 is given by $\alpha_0 = E\left[\frac{1}{F_t(\beta_0)}\right]$. It is of interest to remark that if the $(b(\Delta_t))_t$ are observed, we find as in the MS-ARCH(1) case, $\alpha_0 = E[a(\Delta_t)]$. Indeed, we remark that $a(\Delta_t) F_t(b(\Delta_t)) = 1$ almost surely and the result follows.

We define, for $j \in \mathbb{N}$ and $k \in \{1, \cdots, d\}$, the function $\tilde{a}^{(j)}: \ k \mapsto E\left[\frac{1}{\left(a(k)u_1^2+b(k)\right)^j}\right]$ and with this notation we define $\Theta^{(j)} = \left\{\theta \in \Theta_\gamma, \ \beta^j \rho(\mathbb{P}_{\tilde{a}^{(j)}}) < 1\right\}$. For $i, j \in \{2, 3\}$, let

$$D_t^{\theta_i,\theta_j} = \frac{\partial v_t}{\partial\theta_i\partial\theta_j}(\alpha_0,\beta_0)\left(1-\frac{u_t^2}{v_t(\alpha_0,\beta_0)}\right) + \frac{\partial v_t}{\partial\theta_i}(\alpha_0,\beta_0)\frac{\partial v_t}{\partial\theta_j}(\alpha_0,\beta_0)\left(2\frac{u_t^2}{v_t(\alpha_0,\beta_0)}-1\right),$$

and let $C_{\alpha_0,\beta_0} = E\begin{pmatrix} D_1^{\alpha,\alpha} & D_1^{\alpha,\beta} \\ D_1^{\alpha,\beta} & D_1^{\beta,\beta} \end{pmatrix}$.

In order to obtain the asymptotic normality of the estimator, we state the following assumptions.

B4 For any $\omega_0 > 0$, we have $\theta_0 = (\omega_0, \alpha_0, \beta_0) \in \Theta^{(1)} \cap \Theta^{(2)}$.

B5 There exist $\tilde{\theta} = (\tilde{\omega}, \tilde{\alpha}, \tilde{\beta})'$ such that, $\tilde{\theta} \in \bigcap_{j=1}^{+\infty} \Theta^{(j)}$ and $\tilde{\beta} > 1$.

B6 There exists $\eta_1 > 0$ such that $E|u_t|^{4+\eta_1} < +\infty$.

B7 The matrix C_{α_0,β_0} is non singular.

Theorem 4.4.3. *Under Assumptions **B0-B7**, we have*

$$\sqrt{n}\begin{pmatrix} \alpha_n - \alpha_0 \\ \beta_n - \beta_0 \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, C_{\alpha_0,\beta_0}^{-1} A_{\alpha_0,\beta_0} C_{\alpha_0,\beta_0}^{-1}\right), \tag{4.4.13}$$

where $A_{\alpha,\beta}$ is defined by

$$A_{\alpha,\beta} = E\left[\left(1 - \frac{u_1^2}{v_1(\alpha_0,\beta_0)}\right)^2 \frac{1}{v_1}\frac{\partial v_1}{\partial(\alpha,\beta)'}\frac{1}{v_1}\frac{\partial v_1}{\partial(\alpha,\beta)}\right].$$

4.5 Numerical Experiments

In this section, we investigate the numerical properties of the QMLE of a misspecified GARCH$(1,1)$ (or ARCH(1)) model when the DGP is a MS-GARCH model. We illustrate the results of Theorems 4.3.1, 4.3.2, 4.4.1, 4.4.2 and 4.4.3. These simulations will also show that, for some parameterizations, the predictive ability of the misspecified GARCH$(1,1)$ model is comparable to the predictive ability of the true model.

4.5.1 The MS-ARCH(1) case

In a first time, we study the estimation of a ARCH(1) model when the DGP is a MS-ARCH(1) model. In this subsection, the DGP is Model (4.4.1). We consider the case where the Markov chain $(\Delta_t)_t$ can take two different values $(d = 2)$. Thus, the parameters of the DGP are

$$\tau = (w_1, w_2, a_1, a_2, p_1, p_2)',$$

where the transition matrix of the Markov chain $(\Delta_t)_t$ is given by

$$P = \begin{pmatrix} 1-p_1 & p_1 \\ p_2 & 1-p_2 \end{pmatrix}.$$

From Theorem 4.4.1, in the non stationary case (when $\gamma > 0$), we have the convergence of the estimator α_n toward $\alpha_0 = Ea(\Delta_t)$. We use the following parameterization

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}, \quad \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \quad \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix}. \tag{4.5.1}$$

For this parameterization, the top Lyapunov exponent γ is equal to 0.08. Thus, we are in the non stationary explosive case, the estimator α_n converges toward $\alpha_0 = 4$. For different values of n ($n \in \{500, 1000, 5000, 10000\}$), we draw 1000 samples of size n of Model (4.4.1) with parameterization (4.5.1) and we estimate α_n. Then, we compute $Z_n = \frac{\sqrt{n}}{\sqrt{2\alpha_0^2}}(\alpha_n - \alpha_0)$. By Theorem 4.4.1, the distribution of h_n must converge toward a Gaussian law with unit variance as $n \to +\infty$. The smoothed densities of Z_n are given in Figure 4.1.

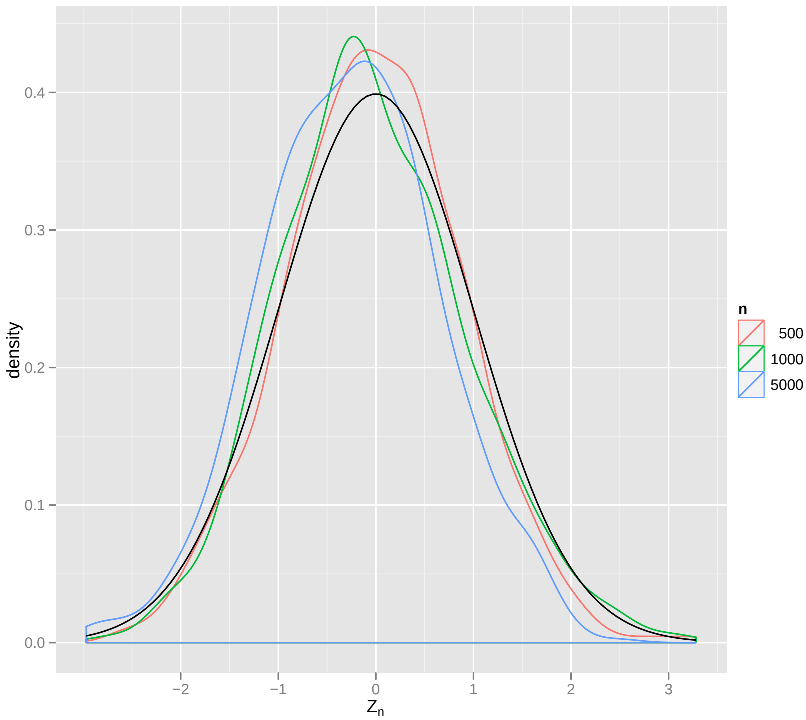


Figure 4.1: Smoothed densities for the estimation of an ARCH(1) model. The curve in bold is the density of the standard Gaussian.

Figure 4.1 illustrates the convergence of the distribution of Z_n toward the standard Gaussian law. Since the model is explosive, greater values of n cannot be computed because the process

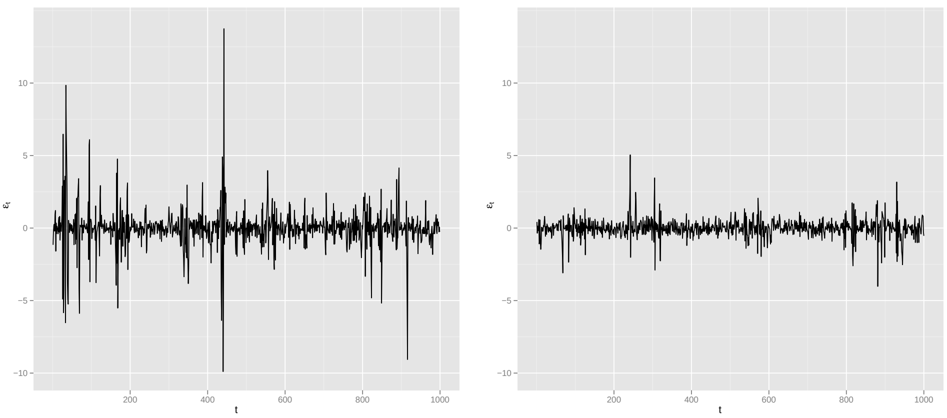
$(\epsilon_t)_t$ would take values that the software would consider as infinity.

In the non explosive case, Theorem 4.3.1 states the convergence of α_n toward a pseudo value α_0. In this case, we do not have necessarily $\alpha_0 = Ea(\Delta_t)$. We study the following parameterizations for different probability specifications of the Markov chain $(\Delta_t)_t$, we consider $p_1 = p_2$ with $p_1 \in \{0.1, 0.05, 0.01\}$.

parameterization	w_1	w_2	a_1	a_2	γ	$\rho(\mathbb{P}_{\tilde{A}^s})$
1	0.1	0.1	1.2	1.6	-0.94	0.96
2	0.1	0.1	0.5	0.9	-1.67	0.93

Table 4.1: parameterizations used in the stationary ARCH case.

In Table 4.1, we give the top Lyapunov exponent associated to the parameterizations and the spectral radius of the matrix $\mathbb{P}_{\tilde{A}^s}$ for $s = 0.05$ and for $p_1 = p_2 = 0.05$. For the other values of p_1 and p_2, $\mathbb{P}_{\tilde{A}^s}$ differs from the value given in Table 4.1 since $\mathbb{P}_{\tilde{A}^s}$ depends on the transition probabilities, but is very close to the value obtained for $p_1 = p_2 = 0.05$. The coefficient γ can be computed using (4.4.2). We can see that these parameterizations satisfy the classical stationary assumption $\gamma < 0$, and the stationary assumption **A0** from Section 4.3 is also statisfied. Note that other values of s could have been chosen, it is only important that there exists one value of s such that **A0** is satisfied. We can also graphically verify the stationarity of the process obtained by these parameterizations by drawing trajectories of the process $(\epsilon_t)_t$ for $p_1 = p_2 = 0.05$, see Figure 4.2.



(a) parameterization #1 (b) parameterization #2

Figure 4.2: Trajectories of $(\epsilon_t)_t$ for different MS-ARCH(1) parameterizations.

For each parameterization, for $n \in \{500, 1000, 5000\}$ and for $p_1 = p_2 \in \{0.1, 0.05, 0.01\}$, we draw 1000 samples of size n and we compute $\theta_n = (\omega_n, \alpha_n)'$. We present the results in Tables 4.2a and 4.2b.

		ω		α	
$p_1 = p_2$	n	Mean	sd	Mean	sd
0.01	500	0.099	0.013	1.409	0.162
0.01	1000	0.099	0.009	1.414	0.118
0.01	5000	0.099	0.004	1.417	0.052
0.05	500	0.100	0.013	1.405	0.144
0.05	1000	0.099	0.009	1.409	0.106
0.05	5000	0.099	0.004	1.413	0.045
0.10	500	0.100	0.014	1.401	0.141
0.10	1000	0.099	0.009	1.407	0.103
0.10	5000	0.099	0.004	1.408	0.043

(a) parameterization #1

		ω		α	
$p_1 = p_2$	N	Mean	sd	Mean	sd
0.01	500	0.099	0.010	0.722	0.137
0.01	1000	0.098	0.007	0.728	0.098
0.01	5000	0.098	0.003	0.733	0.043
0.05	500	0.099	0.010	0.719	0.114
0.05	1000	0.099	0.007	0.724	0.084
0.05	5000	0.098	0.003	0.729	0.035
0.10	500	0.099	0.010	0.715	0.110
0.10	1000	0.099	0.007	0.720	0.081
0.10	5000	0.099	0.003	0.722	0.033

(b) parameterization #2

Table 4.2: Result of the estimation for different MS-ARCH(1) parameterizations.

Since $w_1 = w_2 = 0.1$, it is no surprise to observe that the estimator ω_n appears to converge toward $\omega_0 = 0.1$. The estimator α_n converges toward a pseudo true value α_0. This pseudo true value seems close to $Ea(\Delta_t)$ but contrary to the non stationary case, is not equal to this value. It is also interesting to note that α_0 seems to depend only on the marginal law of $(\Delta_t)_t$. For the different used values of p_1 and p_2, the stationary probabilities of the Markov chain are the same ($\mathbb{P}[\Delta_t = 1] = \mathbb{P}[\Delta_t = 2] = \frac{1}{2}$) and α_0 is also the same for different values of p_1 and p_2 as long as $p_1 = p_2$.

4.5.2 The MS-GARCH$(1,1)$ case

In this subsection, we consider the estimation of a GARCH$(1,1)$ model when the DGP is the MS-GARCH$(1,1)$ model given in (4.4.7). As in the last subsection, we consider the case where the Markov chain $(\Delta_t)_t$ can take values among two states. We study different parameterizations given in Table 4.3 for different values of the transition probabilities of the Markov chain $(\Delta_t)_t$. Note that $\rho(\mathbb{P}_{\tilde{A}^s})$ is calculated for $s = 0.05$ and for different values of p such that $p_1 = p_2 = p$.

Parameterization	w_1	w_2	a_1	a_2	b_1	b_2	γ	$\rho\left(\mathbb{P}_{\tilde{A}^s}\right)$ $p = 0.1$	$\rho\left(\mathbb{P}_{\tilde{A}^s}\right)$ $p = 0.05$	$\rho\left(\mathbb{P}_{\tilde{A}^s}\right)$ $p = 0.01$
1	0.1	1.0	0.1	0.1	0.10	0.80	-0.942	0.961	0.968	0.985
2	0.1	1.0	0.1	0.1	0.70	0.70	-0.236	0.988	0.988	0.988
3	0.1	0.1	0.1	0.1	0.60	0.90	-0.190	0.991	0.991	0.994
4	0.1	0.1	0.1	0.4	0.60	0.60	-0.238	0.988	0.989	0.990
5	0.1	0.1	0.1	0.1	0.20	1.00	-0.595	0.976	0.981	0.996
6	0.1	0.1	0.1	0.1	0.90	1.10	0.084	1.004	1.004	1.005

Table 4.3: MS-GARCH parameterizations.

We study this estimation both in the stationary case (when Assumption **A0** is statisfied, parameterization 1 to 5) and in the non stationary explosive case (when the top Lyapunov exponent γ satisfies $\gamma > 0$, parameterization 6). It is of interest to note that in parameterization 5, regime 1 is stationary with negative top Lyapunov exponent associated to this regime $E\left[\log\left(a_1 u_t^2 + b_1\right)\right]$ and regime 2 is non stationary with positive top Lyapunov exponent. Even with one non stationary regime, the model is stationary. We also remark that the stationarity Assumption **A0** is dependent on the transition probabilities while the top Lyapunov exponent only depends on the stationary probabilities of $(\Delta_t)_t$. For example, using parameterization 5, if $p_1 = p_2 = 0.05$, Assumption **A0** is satisfied with $s = 0.2$ (in this case, we have $\rho\left(\mathbb{P}_{\tilde{A}^s}\right) = 0.975$). But if $p_1 = p_2 = 0.01$, for $s = 0.2$, we have $\rho\left(\mathbb{P}_{\tilde{A}^s}\right) = 1.01$ and Assumption **A0** is not satisfied for this particular value of s. Taking a smaller s, we can show that this parameterization still satisfies Assumption **A0**, for $s = 0.01$, we have $\rho\left(\mathbb{P}_{\tilde{A}^s}\right) = 0.996$.

For each parameterization, for $n \in \{500, 1000, 5000\}$ and for $p_1 = p_2 = 0.05$, we simulate 1000 samples of realizations of Model (4.4.7) of size $n + 500$ and we compute the misspecified QMLE θ_n. Since the GARCH$(1,1)$ model is often considered as a default model and thus used even when the DGP is generated by another model, it is of interest to determine whether the misspecified GARCH$(1,1)$ model is able to partially explain and forecast the dynamic of the observed process $(\epsilon_t)_t$. We evaluate the prediction ability of the GARCH$(1,1)$ model. We have

$$E_{\theta_0,t-1}\left[\epsilon_t^2\right] = E\eta_t^2 E_{\theta_0,t-1}\left[\sigma_t^2(\theta_0)\right] = \omega_0 + \alpha_0\epsilon_{t-1}^2 + \beta_0\sigma_{t-1}^2(\theta_0),$$

where $E_{\theta_0,t-1}$ denotes the expectation under the assumption that the process (ϵ_t) follows Model (4.1.1) and conditionally to $\mathcal{F}_t$, the filtration generated by $\{\epsilon_u,\ u < t\}$. Therefore the best prediction at date $t-1$ for ϵ_t^2 under the assumption that $(\epsilon_t)_t$ is a GARCH$(1,1)$ is

$$\widehat{\epsilon}_{t,t-1}^2 = \omega_0 + \alpha_0\epsilon_{t-1}^2 + \tilde{\sigma}_{t-1}^2(\theta_0).$$

We compare this prediction with the theoretical prediction that we obtain if we use the true model of the DGP and observe the states of the chain $(\Delta_t)_t$. In practice, even if we know that the process is a MS-GARCH model, the states of the chain cannot be observed and we can only obtain the conditional probabilities of presence in each state of the chain. Here we act as if we could observe (Δ_t). Therefore we compare the prediction ability of the GARCH$(1,1)$ model with the best possible prediction if we observe the trajectory $(h_t)_{t\in\{1,\cdots,n\}}$. In order to evaluate the quality of the prevision, we use the Root Mean Squared Errors (RMSE) defined, for the prevision of the misspecified GARCH model by

$$RMSE_1 = \sum_{t=n+1}^{n+500} \left(\epsilon_t^2 - \widehat{\epsilon}^2_{t,t-1}\right)^2.$$

The parameter θ_n is estimated on the 1000 first observations of the process $(\epsilon_t)_t$ (insample) and the RMSE are computed on the 500 next observations (outsample). Let $RMSE_0 = \sum_{t=n+1}^{n+500} \left(\epsilon_t^2 - h_t\right)^2$ be the RMSE of the prevision under the assumption that the processes $(h_t)_t$ and $(\Delta_t)_t$ are observed. We use the ratio $Q = \frac{RMSE_0}{RMSE_1}$ to evaluate the prediction ability of the misspecified GARCH model with respect to the best (and unfeasible) prevision.

Parameterization	n	ω	α	β	Q
1	500	0.071	0.681	0.523	0.603
	1000	0.067	0.673	0.530	0.601
	5000	0.065	0.674	0.530	0.598
2	500	0.147	0.255	0.717	0.900
	1000	0.125	0.252	0.728	0.901
	5000	0.115	0.252	0.731	0.902
3	500	0.059	0.239	0.728	0.918
	1000	0.053	0.236	0.738	0.921
	5000	0.050	0.237	0.740	0.923
4	500	0.086	0.318	0.587	1.000
	1000	0.081	0.319	0.596	1.005
	5000	0.078	0.320	0.601	1.011
5	500	0.046	0.679	0.513	0.636
	1000	0.043	0.678	0.518	0.637
	5000	0.042	0.680	0.518	0.625
6	500	NA	0.312	0.844	0.853
	1000	NA	0.317	0.839	0.874
	5000	NA	0.311	0.829	0.841

Table 4.4: Result of the estimation of a misspecified GARCH$(1, 1)$ model, for $p_1 = p_2 = 0.05$.

In Table 4.4, over the 1000 simulated paths, we present the mean of the estimations of parameters ω_0, α_0 and β_0 and the mean of the statistic Q. All the estimations have been done with $p_1 = p_2 = 0.05$. If we choose another probabilistic specification for the Markov chain, we obtain statistics Q of the same magnitude.

In Parametrizations 2, 3 and 4, there is only one parameter which is affected by the Markov chain (parameter w for Parameterization 2, parameter b for Parameterization 3 and parameter a for Parameterization 4). The ratio Q of the RMSE of the best possible (and unfeasible) prediction and of the RMSE of the prediction obtained with an estimated misspecified GARCH$(1, 1)$ model is, for these three parameterizations above 0.9. When only the parameter a is affected by the changes of Δ_t, the prediction ability of the misspecified model is almost exactly the same of the best possible prediction. The fact that, in this case, the ratio Q is larger than 1 may be surprising and can be due to the fact that, for Parameterization 4, the generated process $(\epsilon_t)_t$ does not possess a finite moment of order 4 and thus, the asymptotic behavior of the RMSEs is undetermined. For these three parameterizations, the DGP possesses a finite moment of order 2, it is thus of interest to remark that, in each case, we have $\alpha_n + \beta_n < 1$. The constraint $\alpha + \beta < 1$ is a sufficient and necessary condition for the existence of a second order stationary solution to the GARCH$(1,1)$ model. In these cases, the misspecified GARCH$(1,1)$ model shares the property of existence of a finite moment of order 2 for $(\epsilon_t)_t$ with the DGP.

The case of Parameterization 1 is more surprising. In this case, the DGP is strictly stationary and second order stationary and none of its regimes are non stationary. The DGP also possesses a finite moment or order 4. It is thus a surprise to observe that the sum of the estimates α_n and β_n is significantly larger than 1. We also remark that, in this case, the prediction ability of the GARCH modeling is much lower than in the other cases.

We also observe $\alpha_n + \beta_n > 1$ for the results of the inference of Parameterization 5. In this case, the DGP does not possess a finite moment of order 2, it is thus logical that the misspecified GARCH$(1,1)$ model satisfies this relation. In this case, the two regimes possess a very different behavior. In Regime 1, there is very few persistence of volatility whereas Regime 2 is explosive. In consequence, the misspecified GARCH$(1,1)$ model is unable to explain the dynamic of the process as well as in Parameterizations 2, 3 or 4.

Parameterization 6 corresponds to the non stationary case of Section 4.4. The parameter ω cannot be inferred since the model is explosive. The optimization algorithm will only return the initial value given for ω. We remark that, in this case, the misspecified GARCH model is able to predict a large fraction of the dynamic of the observed process.

Hillebrand (2005) showed that the inference of a GARCH$(1,1)$ model when the DGP presents parameter changes gives estimates satisfying $\alpha_n + \beta_n \approx 1$. They exhibit a spurious IGARCH convergence. In his paper, the DGP is not a MS-GARCH but a non stationary GARCH model with a single structural break in the middle of the samples. When the transition probabilities p_1 and p_2 become smaller, the MS-GARCH$(1,1)$ model looks more similar to the GARCH model with a single structural change in the parameters. Figure 4.3 shows that we obtain a $\alpha_n + \beta_n$ close to 1 when the probability of remaining in the same state is high.

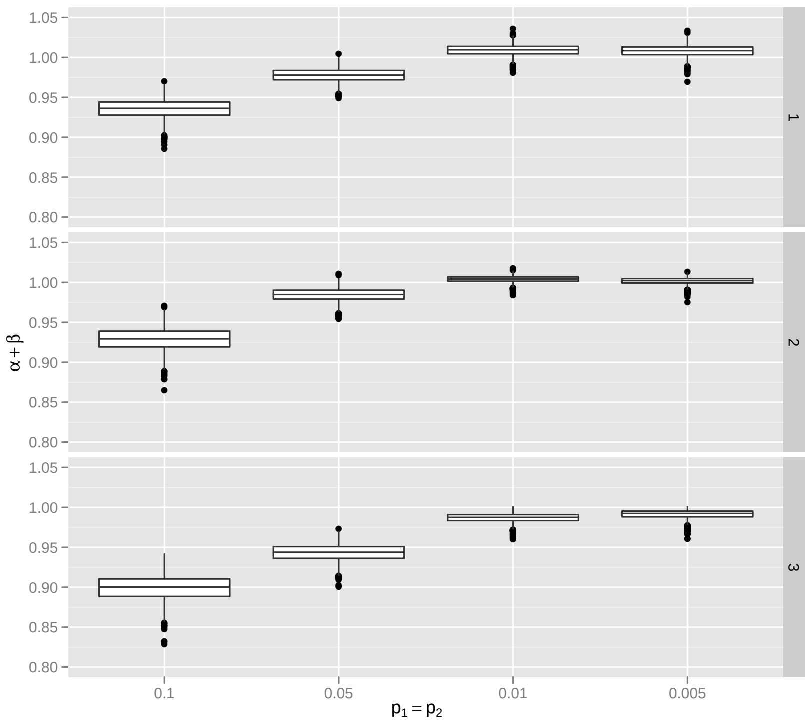


Figure 4.3: Box-plots of the estimation of $\alpha + \beta$ for Parameterization 1 (top panel), Parameterization 2 (middle panel) and Parameterization 3 (bottom panel) and for $n = 5000$

From this last property, we can also conclude that, contrarily to the MS-ARCH(1) case, the estimates of the GARCH$(1, 1)$ model when the DGP is a MS-GARCH$(1, 1)$ process does not only depend on the marginal distribution of the Markov chain $(\Delta_t)_t$ but depends on the transition probabilities of $(\Delta_t)_t$.

4.6 Proofs

In all the proofs, $K > 0$ and $\rho \in (0, 1)$ denote generic constants whose value can change throughout the proof.

4.6.1 Proof of Lemma 4.2.1

We define, for $n \in \mathbb{N}^*$

$$\underline{z}_{t,n} = \underline{b}_t + \sum_{k=1}^{n} A_t \cdots A_{t-k+1} \underline{b}_{t-k}, \quad \underline{z}_{t,\infty} = \underline{b}_t + \sum_{k=1}^{+\infty} A_t \cdots A_{t-k+1} \underline{b}_{t-k}. \tag{4.6.1}$$

Since all the terms of the matrices and vectors $(A_t)_t$ and $(\underline{b}_t)_t$ are almost surely in $\mathbb{R}_+ \cup \{+\infty\}$, the quantity $\underline{z}_{t,\infty}$ is well defined in $\mathbb{R}_+ \cup \{+\infty\}$ and $\underline{z}_{t,n}$ converges toward $\underline{z}_{t,\infty}$ when n tends to infinity.

We use the matrix norm defined by $\|A\| = \sum_{i,j} |A(i,j)|$. Since $s < 1$, using $\left(\sum_i u_i\right)^s \leq \sum_i u_i^s$ for any sequence of positive numbers (u_i), we have for any matrix A and B, $\|AB\|^s \leq \left\|A^{(s)}B^{(s)}\right\|$. For $k \in \mathbb{N}^*$, we have

$$\begin{aligned}
E\|A_t \cdots A_{t-k+1}\underline{b}_{t-k}\|^s &\leq E\left[\|A_t \cdots A_{t-k+1}\|^s \|\underline{b}_{t-k}\|^s\right] \\
&\leq KE\left[\|A_t^{(s)} \cdots A_{t-k+1}^{(s)}\|\right] E\|\underline{\bar{b}}_{t-k}\|^s \\
&\leq K\left\|E\left[A_t^{(s)} \cdots A_{t-k+1}^{(s)}\right]\right\| \\
&\leq K\left\|E\left[E\left[A_t^{(s)} \cdots A_{t-k+1}^{(s)} | \Delta_t, \cdots, \Delta_{t-k+1}\right]\right]\right\| \\
&\leq K\left\|E\left[\tilde{A}^s(\Delta_t) \cdots \tilde{A}^s(\Delta_{t-k+1})\right]\right\|, \qquad (4.6.2)
\end{aligned}$$

where $\underline{\bar{b}}_t = \left(\max_{k\in\{1,\cdots,d\}} w(k)u_t^2, 0, \cdots, \max_{k\in\{1,\cdots,d\}} w(k), 0, \cdots\right)'$. The quantity $\underline{\bar{b}}_{t-k}$ does not depend on the Markov chain (Δ_t) and is thus independent of the sequence $(A_t)_{t \geq t-k+1}$. We used the fact that $(u_t)_t$ has finite variance to obtain $E\|\underline{\bar{b}}_t\|^s < +\infty$. The inversion of norm and expectation can be done since all the terms of the different matrices are positive. Finally, we used the fact that the sequence $(A_{t-i})_{i\in\{0,\cdots,k-1\}}$ is independent conditionally to $(\Delta_{t-i})_{i\in\{0,\cdots,k-1\}}$.

Now, using Lemma 1 in Francq and Zakoïan (2005), we obtain

$$\left\|E\left[\tilde{A}^s(\Delta_t) \cdots \tilde{A}^s(\Delta_{t-k+1})\right]\right\| = \left\|\mathbb{I}\mathbb{P}^k_{\tilde{A}^s}\Pi_1\right\| \leq K\left\|\mathbb{P}^k_{\tilde{A}^s}\right\|.$$

By Assumption **A0**, it follows from the Jordan decomposition that $\left\|\mathbb{P}^k_{\tilde{A}^s}\right\|$ converges to zero at an exponential rate. Therefore, from (4.6.2), we obtain

$$\sum_{k=1}^{+\infty} E\|A_t \cdots A_{t-k+1}\underline{b}_{t-k}\|^s < +\infty.$$

In view of the monotone convergence theorem, we obtain $E\|\underline{z}_{t,\infty}\|^s < +\infty$, therefore we have $\underline{z}_{t,\infty}(q+1) \in \mathbb{R}^+$ with probability 1. Defining $\epsilon_t = \sqrt{\underline{z}_{t,\infty}(q+1)}u_t$, we define a solution of Model (4.2.1). This solution is non anticipative, strictly stationary and ergodic (see Billingsley (1995), Theorem 36.4) and satisfies (4.2.6).

Let $\underline{z}_t$ be the stationary solution of (4.2.5) defined in (4.6.1) and let $\underline{z}_t^*$ be another strictly stationary solution. We assume that $\mathbb{P}\left[\|\underline{z}_t - \underline{z}_t^*\| \neq 0\right] > 0$. We have

$$\|\underline{z}_t - \underline{z}_t^*\| = |A_t(\underline{z}_{t-1} - \underline{z}_{t-1}^*)| \leq \|A_t \cdots A_{t-n+1}\| \; \|\underline{z}_{t-n} - \underline{z}_{t-n}^*\|.$$

The series $\sum_{k=1}^{+\infty} A_t \cdots A_{t-k+1}w_{t-k}$ converges, thus we have $\lim_{k\to+\infty} \|A_t \cdots A_{t-k+1}\| = 0$. Consequently, $\mathbb{P}\left[\lim_{n\to+\infty} \|\underline{z}_{t-n} - \underline{z}_{t-n}^*\| = +\infty\right] > 0$, which entails that either $\limsup_{n\to+\infty} \|\underline{z}_{t-n}\| = +\infty$

or $\limsup_{n\to+\infty} \|\underline{z}^*_{t-n}\| = +\infty$ with positive probability. This is in contradiction with the stationary assumption for $\underline{z}_t$ and $\underline{z}^*_t$. Therefore the solution of (4.2.1) is unique.

4.6.2 Proof of Remark 4.3.1

We prove that in the case $d = 2$, we have $Sp(\mathbb{P}_B) = Sp(\mathbb{Q}_B)$. We define

$$\tilde{B} = \begin{pmatrix} B(1) & \cdots & 0 \\ 0 & \ddots & \vdots \\ 0 & \cdots & B(d) \end{pmatrix},$$

and we have

$$\mathbb{P}_B = \tilde{B}\left(\mathbb{P} \otimes I_q\right).$$

Using the Bayes formula, we obtain $q(i,j) = p(j,i)\frac{\Pi(j)}{\Pi(i)}$. Then we remark that

$$\mathbb{Q}_B = \text{diag}\left(\Pi(1), \cdots, \Pi(d)\right) {}^t(\mathbb{P})_B \text{diag}\left(\frac{1}{\Pi(1)}, \cdots, \frac{1}{\Pi(d)}\right).$$

Therefore we have $Sp(\mathbb{Q}_B) = Sp({}^t(\mathbb{P})_B)$, where for a generic matrix M, $Sp(M)$ is the spectrum of M. We have

$${}^t(\mathbb{P})_B = \tilde{B}\left({}^t\mathbb{P} \otimes I_q\right).$$

If $d = 2$ and if the matrices $B(1)$ and $B(2)$ are invertible, we have

$$\begin{aligned}
\lambda \in Sp(\mathbb{P}_B) &\Leftrightarrow \det\Big(\lambda I_{2q} - \tilde{B}\left(\mathbb{P} \otimes I_q\right)\Big) = 0 \Leftrightarrow \det\Big(\lambda \tilde{B}^{-1} - \mathbb{P} \otimes I_q\Big) = 0 \\
&\Leftrightarrow \det\begin{pmatrix} \lambda B(1)^{-1} + p(1,1)I_q & p(1,2)I_q \\ p(2,1)I_q & \lambda B(2)^{-1} + p(2,2)I_q \end{pmatrix} = 0 \\
&\Leftrightarrow \det\left\{\left(\lambda B(1)^{-1} + p(1,1)I_q\right)\left(\lambda B(2)^{-1} + p(2,2)I_q\right) - p(1,2)p(2,1)I_q\right\} = 0 \\
&\Leftrightarrow \det\begin{pmatrix} \lambda B(1)^{-1} + p(1,1)I_q & p(2,1)I_q \\ p(1,2)I_q & \lambda B(2)^{-1} + p(2,2)I_q \end{pmatrix} = 0 \\
&\Leftrightarrow \det\Big(\lambda \tilde{B}^{-1} - {}^t\mathbb{P} \otimes I_q\Big) = 0 \\
&\Leftrightarrow \det\Big(\lambda I_{2q} - \tilde{B}\left({}^t\mathbb{P} \otimes I_q\right)\Big) = 0 \Leftrightarrow \lambda \in Sp(({}^t\mathbb{P})_B) \Leftrightarrow \lambda \in Sp(\mathbb{Q}_B).
\end{aligned}$$

4.6.3 Proof of Proposition 4.3.1

Let $\theta^* \in \Theta^*$, we prove that $El_t(\theta^*) < +\infty$. For that, we write the DGP in another vectorial form. We define $\underline{h}_t = (h_t, \cdots, h_{t-p+1})'$ and we have

$$\underline{h}_t = \underline{c}_t + B_t \underline{h}_{t-1},$$

with $\underline{c}_t = \left(w(\Delta_t) + \sum_{i=1}^q a_i(\Delta_t)\epsilon_{t-i}^2, 0, \cdots, 0\right)' \in \mathbb{R}^q$. We have

$$\underline{h}_t = \underline{c}_t + \sum_{k=1}^n B_t \cdots B_{t-k+1}\underline{c}_{t-k} + B_t \cdots B_{t-n+1}{}_{t-n}.$$

We define, for $k \in \{1, \cdots, d\}$, $\tilde{c}^{(s)}(k) = E\left[\underline{c}_t^{(s)} \mid \Delta_t = k\right]$. Using once again Lemma 1 in Francq and Zakoïan (2005), the existence of a moment of order $2s$ for the process $(\epsilon_t)_t$ and similar arguments as in the proof of Lemma 4.2.1, we obtain

$$\begin{aligned}
E\left[\left\|B_t \cdots B_{t-k+1}\underline{c}_{t-k}\right\|^s\right] &\le \left\|E\left[B_t^{(s)} \cdots B_{t-k+1}^{(s)}\underline{c}_{t-k}^{(s)}\right]\right\| \\
&\le \left\|E\left[B(\Delta_t)^{(s)} \cdots B(\Delta_{t-k+1})^{(s)}\tilde{c}^{(s)}(\Delta_{t-k})\right]\right\| \\
&\le \left\|\mathbb{I}\mathbb{P}_{B^{(s)}}^k \Pi_{\tilde{c}^{(s)}}\right\| \le K\left\|\mathbb{P}_{\tilde{A}^s}^k\right\|.
\end{aligned}$$

Using Assumption **A0**, the convergence of the series $\sum_{k=1}^{+\infty} E\left[\left\|B_t \cdots B_{t-k+1}\underline{c}_{t-k}\right\|^s\right]$ follows. This also yields

$$\underline{h}_t = \sum_{k=0}^{+\infty} B_t \cdots B_{t-k+1}\underline{c}_{t-k} < +\infty, \text{ a.s.},$$

with the convention that $B_t \cdots B_{t-k+1} = 1$ when $k = 0$.

We have $l_t(\theta) = \log \sigma_t^2 + \frac{\epsilon_t^2}{\sigma_t^2} = \log \sigma_t^2 + u_t^2 \frac{h_t^2}{\sigma_t^2}$. With Lemma 4.2.1, we have $E|\epsilon_t|^{2s} < +\infty$ and consequently $E\left|\omega + \alpha\epsilon_{t-k-1}^2\right|^s < +\infty$. Thus, from Assumption **A1**, we obtain $E|\sigma_t|^{2s} < +\infty$ and consequently

$$E \log \sigma_t^2 < +\infty. \tag{4.6.3}$$

Then we focus on the term $\frac{h_t}{\sigma_t^2}$, we have

$$\begin{aligned}
\frac{h_t}{\sigma_t^2} &= \frac{1}{\sigma_t^2}\sum_{k=0}^{+\infty}\{B_t \cdots B_{t-k+1}\}(1,1)w(\Delta_{t-k}) \\
&\quad + \frac{1}{\sigma_t^2}\sum_{k=0}^{+\infty}\{B_t \cdots B_{t-k+1}\}(1,1)\left\{\sum_{i=1}^q \alpha_i(\Delta_{t-k})\epsilon_{t-k-i}^2\right\}.
\end{aligned} \tag{4.6.4}$$

With Assumption **A1**, there exists $\underline{w} > 0$ such that for any $\theta \in \Theta$, we have $w > \underline{w}$. In consequence we also have for any $\theta \in \Theta$, $\sigma_t^2 > \underline{w}$. We begin by treating the first term of (4.6.4), we have

$$\frac{1}{\sigma_t^2}\sum_{k=0}^{+\infty}\{B_t \cdots B_{t-k+1}\}(1,1)w(\Delta_{t-k}) \le K\sum_{k=0}^{+\infty}\|B_t \cdots B_{t-k+1}\|.$$

With Assumption **A0** and some already used arguments, we obtain

$$E\left[\frac{1}{\sigma_t^2}\sum_{k=0}^{+\infty}\{B_t\cdots B_{t-k+1}\}(1,1)w(\Delta_{t-k})\right] < +\infty. \tag{4.6.5}$$

About the second term in (4.6.4), we use the fact that for any $\delta > 0$ and for $x \geq 0$, we have $\frac{x}{1+x} \leq x^{\delta}$. Therefore, for any $r > 0$ we have

$$\begin{aligned}
C_t(\theta) &= \frac{1}{\sigma_t^2}\sum_{k=0}^{+\infty}\{B_t\cdots B_{t-k+1}\}(1,1)\left\{\sum_{i=1}^{q}\alpha_i(\Delta_{t-k})\epsilon_{t-k-i}^2\right\} \\
&\leq K\sum_{i=1}^{q}\sum_{k=0}^{+\infty}\|B_t\cdots B_{t-k+1}\|\frac{\epsilon_{t-k-i}^2}{\sigma_t^2} \\
&\leq K\sum_{i=1}^{q}\sum_{k=0}^{+\infty}\|B_t\cdots B_{t-k+1}\|\frac{\epsilon_{t-k-i}^2}{\omega+\beta^{k+i-1}\epsilon_{t-k-i}^2} \\
&\leq K\sum_{i=1}^{q}\sum_{k=0}^{+\infty}\frac{\|B_t\cdots B_{t-k+1}\|}{\beta^k}\beta^{(i-1)(s-1)}\beta^{rk}\epsilon_{t-k-i}^{2s}.
\end{aligned}$$

The main issue is that ϵ_t depends of any realization of the process $(\Delta_t)_t$. At this point, we need a modified version of Lemma 1 in Francq and Zakoïan (2005), which can be easily proved.

Lemma 4.6.1. *Let $f:\ \{1,\cdots,d\}\to\mathcal{M}_{n\times n}(\mathbb{R})$ and $g:\ \{1,\cdots,d\}\to\mathcal{M}_{n\times n'}(\mathbb{R})$. We define $\underline{\Delta}_t = \{\Delta_{t'},\ t'\leq t\}$ and for $k>0$ and $h>k$, $g_{h,k}(i) = E\left[g(\underline{\Delta}_{t-h})|\Delta_{t-k+1}=i\right]$, we have*

$$E\left[f(\Delta_t)\cdots f(\Delta_{t-k+1})g(\underline{\Delta}_{t-h})\right] = \left(g_{h,k}(1),\cdots,g_{h,k}(d)\right)\mathbb{Q}_f^2\Pi_1. \tag{4.6.6}$$

Since all the components of the matrices $(B_t)_t$ are positive, we have

$$E\left[\|B_t\cdots B_{t-k+1}\|\,\epsilon_{t-k-i}^{2s}\right] = \left\|E\left[B_t\cdots B_{t-k+1}\left\{\epsilon_{t-k-i}^{2s}I_q\right\}\right]\right\|.$$

Besides, we have for any $k\geq 0$, $E\left[B_{t-k}|\ \underline{\Delta}_t\right] = B_{t-k} = B(\Delta_{t-k})$, and $E\left[\epsilon_{t-k-i}^{2r}|\ \underline{\Delta}_t\right] = E\left[\epsilon_{t-k-i}^{2r}|\ \underline{\Delta}_{t-k-i}\right]$. Therefore, with Lemma 4.6.1, we have

$$\begin{aligned}
E\left[B_t\cdots B_{t-k+1}\epsilon_{t-k-i}^{2s}\right] &= E\left[B(\Delta_t)\cdots B(\Delta_{t-k+1})\left\{E\left[\epsilon_{t-k-i}^{2s}|\ \underline{\Delta}_{t-k-i}\right]I_q\right\}\right] \\
&= \left(g_{i,k}(1),\cdots,g_{i,k}(d)\right)\mathbb{Q}_B^k\Pi_1
\end{aligned}$$

where $g_{i,k}(j) = E\left[E\left[\epsilon_{t-k-i}^{2s}|\ \underline{\Delta}_{t-k-i}\right]|\Delta_{t-k+1}=j\right]I_q$. Therefore, with the existence of a moment of order $2s$ for the process (ϵ_t), we obtain

$$E\left[\|B_t \cdots B_{t-k+1}\| \, \epsilon_{t-k-i}^{2s}\right] \leq K\|\mathbb{Q}_B^k\|.$$

We finally obtain

$$EC_t(\theta^*) \leq K \sum_{i=1}^{q} \sum_{k=0}^{+\infty} \beta^{*sk} \frac{\|\mathbb{Q}_B^k\|}{\beta^{*k}}.$$

Using Assumption **A2** and **A3**, we obtain that $\frac{\|\mathbb{Q}_B^k\|}{\beta^{*k}}$ converges to zero when $k \to +\infty$ and therefore, since $\beta^* < 1$, we have $EC_t(\theta^*) < +\infty$ and consequently

$$E\frac{h_t}{\sigma_t^2(\theta^*)} < +\infty. \tag{4.6.7}$$

With (4.6.3), (4.6.5) and (4.6.7), we obtain (4.3.2). Using Assumption **A4** and the existence of a moment of order $2s$, we obtain (4.3.3).

4.6.4 Proof of Theorem 4.3.1

This theorem can be proved as Theorem 2.1 of Francq and Zakoïan (2004), using Assumption **A4** and Proposition 4.3.1, the proof is classical and straightforward.

4.6.5 Proof of Theorem 4.3.2

Lemma 4.6.2. *Under Assumptions of Theorem 4.3.2, there exists a neighborhood of θ_0, $V_1(\theta_0)$ such that*

$$E\left[\sup_{\theta \in V(\theta_0)} \frac{h_t}{\sigma_t^2(\theta)}\right] < +\infty. \tag{4.6.8}$$

Proof. For any $\theta \in \Theta$, we have

$$\frac{h_t}{\sigma_t^2(\theta)} = \frac{h_t}{\sigma_t^2(\theta_0)} \times \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}. \tag{4.6.9}$$

For any $y > 0$ and for $V(\theta_0)$ a neighborhood of θ_0, we have

$$\left\|\sup_{\theta \in V(\theta_0)} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\right\|_y \leq \sup_{\theta \in V(\theta_0)} \frac{1}{\sigma_t^2(\theta)} K \sum_{k=0}^{+\infty} \beta_0^k + \left\|\sum_{k=0}^{+\infty} \alpha_0 \beta_0^k \sup_{\theta \in V(\theta_0)} \frac{\epsilon_{t-k-1}^2}{w + \beta^k \epsilon_{t-k-1}^2}\right\|_y.$$

For any $r \in (0,1]$ and for $x > 0$, we have $x/(1+x) < x^r$. We take r such that $r * y \leq s$. Now, defining $V_1(\theta_0) = \left\{\theta \in \Theta,\ \|\theta_0 - \theta\| < 1 - \beta_0^{\frac{r}{1-r}}\right\}$, we obtain

$$\left\| \sup_{\theta \in V_1(\theta_0)} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} \right\|_y < +\infty. \tag{4.6.10}$$

Now, using Assumption **A5**, (4.6.9) and the Hölder inequality, for $y = \frac{1+\nu}{\nu}$, we have obtained the existence of a neighborhood $V(\theta_0)$ of θ_0 such that

$$E\left[\sup_{\theta \in V_1(\theta_0)} \frac{h_t}{\sigma_t^2(\theta)}\right] \leq \left\| \frac{h_t}{\sigma_t^2(\theta_0)} \right\|_{1+\nu} \left\| \sup_{\theta \in V_1(\theta_0)} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} \right\|_y < +\infty.$$

□

Lemma 4.6.3. *Under the assumptions of Theorem 4.3.2, we have*

$$E\left[\frac{\partial l_t}{\partial \theta}(\theta_0)\right] = 0. \tag{4.6.11}$$

Proof. We have, for any $\theta \in \Theta$

$$\frac{\partial l_t}{\partial \theta} = \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \left(1 - \frac{h_t}{\sigma_t^2} u_t^2\right). \tag{4.6.12}$$

The function $\theta \mapsto El_t(\theta)$ is continuous and by Assumption **A4** satisfies $\frac{\partial}{\partial \theta} El_t(\theta_0) = 0$. Therefore it is sufficient to prove

$$E\frac{\partial}{\partial \theta} l_t(\theta_0) = \frac{\partial}{\partial \theta} El_t(\theta_0).$$

For $\theta \in \Theta$ and from (4.3.1), we have

$$\frac{\partial \sigma_t^2}{\partial w} = \frac{1}{1-\beta} \tag{4.6.13}$$

$$\frac{\partial \sigma_t^2}{\partial \alpha} = \sum_{k=0}^{+\infty} \beta^k \epsilon_{t-k-1}^2 \tag{4.6.14}$$

$$\frac{\partial \sigma_t^2}{\partial \beta} = \sum_{k=0}^{+\infty} k\beta^{k-1}(\omega + \alpha\epsilon_{t-k+1}^2). \tag{4.6.15}$$

We define $\delta > 0$ as the solution of

$$\frac{1+\delta}{(1-\delta)^{1-r}} = \frac{1}{\beta_0^r}. \tag{4.6.16}$$

This equation has a solution. Indeed, the function $\delta \mapsto \frac{1+\delta}{(1-\delta)^{1-r}}$ is strictly increasing on $\mathbb{R}_+$ and is equal to 1 when $\delta = 0$ and tends to infinity when δ tends to 1, therefore since $1/\beta_0^r > 1$, the solution exists and is unique. Thus, for any $0 < \iota < \delta$, we have $\frac{1+\iota}{(1-\iota)^{1-r}} < \frac{1}{\beta_0^r}$.

Let $V_2(\theta_0) \subset \Theta$ be a neighborhood of θ_0 such that, for any $\theta = (w, \alpha, \beta)' \in V_2(\theta_0)$ we have $(1-\delta)\beta_0 \leq \beta \leq (1+\delta)\beta_0$. Let $\tilde{\theta} = (\tilde{w}, \tilde{\alpha}, \tilde{\beta}) \in V_2(\theta_0)$ and $\hat{\theta} = (\hat{w}, \hat{\alpha}, \hat{\beta}) \in V_2(\theta_0)$ be such that $\tilde{\beta} = (1-\delta)\beta_0,\ \hat{\beta} = (1+\delta)\beta_0$ and for $i \in \{1, \cdots, 2\},\ \tilde{\theta}_i < \theta_{0i} < \hat{\theta}_i$. We recall that $V_1(\theta_0)$ is the neighborhood of θ_0 defined in Lemma 4.6.2, we define $V_3(\theta_0) = V_1(\theta_0) \cap V_2(\theta_0)$. We have, for any $i \in \{1, \cdots, 3\}$ and for any $\theta \in V_3(\theta_0)$

$$\left|\frac{\partial}{\partial \theta_i} l_t(\theta)\right| \leq \left|\frac{1}{\sigma_t^2(\tilde{\theta})} \frac{\partial \sigma_t^2}{\partial \theta_i}(\hat{\theta})\right| \left|1 + u_t^2 \frac{h_t}{\sigma_t^2(\tilde{\theta})}\right|. \tag{4.6.17}$$

We prove that the right term of the last equation is integrable. We have using the Hölder inequality

$$\left\|1 + u_t^2 \frac{h_t}{\sigma_t^2(\tilde{\theta})}\right\|_{1+\eta_2/2} \leq 1 + K \left\|\frac{h_t}{\sigma_t^2(\theta_0)}\right\|_{1+\eta_2} \left\|\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\tilde{\theta})}\right\|_{2\frac{1+\eta_2}{\eta_2}}.$$

Using Assumption **A5** and already used arguments, we obtain

$$\left\|1 + u_t^2 \frac{h_t}{\sigma_t^2(\tilde{\theta})}\right\|_{1+\eta_2/2} < +\infty. \tag{4.6.18}$$

With (4.6.13), we have $\left\|\frac{1}{\sigma_t^2(\tilde{\theta})} \frac{\partial \sigma_t^2}{\partial w}(\hat{\theta})\right\|_{\frac{2+\eta_2}{\eta_2}} < +\infty$. From (4.6.14), we obtain, with $r < \frac{s\eta_2}{2+\eta_2}$

$$\begin{aligned}
\left\|\frac{1}{\sigma_t^2(\tilde{\theta})} \frac{\partial \sigma_t^2}{\partial \alpha}(\hat{\theta})\right\|_{\frac{2+\eta_2}{\eta_2}} &\leq \left\|\sum_{k=0}^{+\infty} \hat{\beta}^k \frac{\epsilon_{t-k-1}^2}{\tilde{w} + \tilde{\beta}^k \epsilon_{t-k-1}^2}\right\|_{\frac{2+\eta_2}{\eta_2}} \\
&\leq K \sum_{k=0}^{+\infty} \frac{\hat{\beta}^k}{\tilde{\beta}^k} \tilde{\beta}^{kr} \left\| |\epsilon_{t-k-1}|^{2r} \right\|_{\frac{2+\eta_2}{\eta_2}} \\
&< +\infty.
\end{aligned}$$

We use the fact that from (4.6.16), we have $\hat{\beta}\tilde{\beta}^{r-1} < 1$. With similar arguments, we obtain

$$\left\|\frac{1}{\sigma_t^2(\tilde{\theta})}\frac{\partial\sigma_t^2}{\partial\beta}(\hat{\theta})\right\|_{\frac{2+\eta_2}{\eta_2}} < +\infty$$

Finally with the Hölder inequality, (4.6.17) and (4.6.18), we obtain for $i \in \{1, \cdots, 3\}$

$$\sup_{\theta\in V_3(\theta_0)}\left|\frac{\partial}{\partial\theta_i}l_t(\theta)\right| \leq \left|\frac{1}{\sigma_t^2(\tilde{\theta})}\frac{\partial\sigma_t^2}{\partial\theta_i}(\hat{\theta})\right|\left|1+u_t^2\frac{h_t}{\sigma_t^2(\tilde{\theta})}\right|,$$

and

$$E\left[\left|\frac{1}{\sigma_t^2(\tilde{\theta})}\frac{\partial\sigma_t^2}{\partial\theta_i}(\hat{\theta})\right|\left|1+u_t^2\frac{h_t}{\sigma_t^2(\tilde{\theta})}\right|\right] < +\infty.$$

We conclude with the theorem of derivation under the integral sign and obtain (4.6.11). □

Lemma 4.6.4. *Under the assumptions of Theorem 4.3.2, we have for Θ^* a compact set which contains θ_0 and which is contained is the interior of Θ, for $(i, j, k) \in \{1, \cdots, 3\}$ and for any $y > 0$*

$$E\sup_{\theta\in\Theta^*}|\phi_{t,ij}|^y < +\infty, \quad E\sup_{\theta\in\Theta^*}|\phi_{t,i}|^y < +\infty, \quad E\sup_{\theta\in\Theta^*}|\phi_{t,ijk}|^y < +\infty, \tag{4.6.19}$$

where

$$\phi_{t,i} = \phi_{t,i}(\theta) = \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}, \quad \phi_{t,ij} = \frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}, \quad \phi_{t,ijk} = \frac{1}{\sigma_t^2}\frac{\partial^3\sigma_t^2}{\partial\theta_i\partial\theta_j\partial\theta_k}.$$

Proof. This result can be proven juste as in the well specified GARCH case, see Francq and Zakoïan (2004). □

Lemma 4.6.5. *Under the assumptions of Theorem 4.3.2, we have*

$$E\left\|\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}\right\| < +\infty, \quad E\left\|\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right\| < +\infty. \tag{4.6.20}$$

Besides, there exists a neighborhood $V(\theta_0) \subset \Theta$ of θ_0 such that, for $i, j, k \in \{1, 2, 3\}$

$$E\sup_{\theta\in V(\theta_0)}\left|\frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| + \infty. \tag{4.6.21}$$

Proof. We have for $\theta \in \Theta$ and for $i, j \in \{1, \cdots, 3\}$

$$\begin{aligned}
\frac{\partial l_t(\theta)}{\partial \theta_i} &= \phi_{t,i} \left(1 - u_t^2 \frac{h_t}{\sigma_t^2}\right) \\
\frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j} &= \phi_{t,ij} \left(1 - u_t^2 \frac{h_t}{\sigma_t^2}\right) + \phi_{t,i}\phi_{t,j} \left(2u_t^2 \frac{h_t}{\sigma_t^2} - 1\right).
\end{aligned}$$

With Lemma 4.6.4, Assumption **A6**, the Hölder inequality and similar arguments as those used to obtain (4.6.18), we obtain (4.6.20). Writing

$$\begin{aligned}
\frac{\partial^3 l_t}{\partial \theta_i \partial \theta_j \partial \theta_k} \quad = \quad & \left(2u_t^2 \frac{h_t}{\sigma_t^2} - 1\right) (\phi_{t,i}\phi_{t,jk} + \phi_{t,j}\phi_{t,ik} + \phi_{t,k}\phi_{t,ik}) + \left(1 - u_t^2 \frac{h_t}{\sigma_t^2}\right) \phi_{t,ijk} \\
& + \left(2 - 6u_t^2 \frac{h_t}{\sigma_t^2}\right) \phi_{t,i}\phi_{t,j}\phi_{t,k}, \qquad (4.6.22)
\end{aligned}$$

we obtain with already used arguments (4.6.21). □

Lemma 4.6.6. *Under the assumptions of Theorem 4.3.2, the matrix $B(\theta_0)$ is positive definite.*

Proof. We have

$$\frac{\partial l_t(\theta)}{\partial \theta} \frac{\partial l_t(\theta)}{\partial \theta'} = (1 - u_t^2 \frac{h_t}{\sigma_t^2})^2 \phi_t \phi_t',$$

where $\phi_t = (\phi_{t,i})_{i \in \{1, \cdots, 3\}}$. We prove that $B(\theta_0)$ is definite. If there exists a non zero vector $\lambda \in \mathbb{R}^3$ such that $\lambda' A(\theta_0) \lambda = 0$, then we have

$$E \left(\lambda' \frac{\partial l_t(\theta_0)}{\partial \theta}\right)^2 = 0,$$

which implies $\left(1 - u_t^2 \frac{h_t}{\sigma_t^2}\right) \lambda' \phi_t = 0$ almost surely. Let $Z_{t-1} = \lambda' \phi_t$, Z_{t-1} is independent of u_t and

$Z_{t-1} = 0$ implies that $u_t^2 = \frac{\sigma_t^2}{h_t}$ or $|u_t| = \frac{\sigma_t}{\sqrt{h_t}}$. If $\mathbb{P}[Z_{t-1} \neq 0] > 0$, we have

$$\begin{aligned}
\mathrm{Var}[|u_t|] &= \mathrm{Var}\left[|u_t||Z_{t-1} \neq 0\right] = \mathrm{Cov}\left(|u_t|, \frac{\sigma_t}{\sqrt{h_t}}|Z_{t-1} \neq 0\right) \\
&= E\left[|u_t|\frac{\sigma_t}{\sqrt{h_t}}|Z_{t-1} \neq 0\right] - E|u_t|E\left[\frac{\sigma_t}{\sqrt{h_t}}|Z_{t-1} \neq 0\right] \\
&= E\left[E\left[|u_t|\frac{\sigma_t}{\sqrt{h_t}}|Z_{t-1} \neq 0, \frac{\sigma_t}{\sqrt{h_t}}\right]|Z_{t-1} \neq 0\right] - E|u_t|E\left[\frac{\sigma_t}{\sqrt{h_t}}|Z_{t-1} \neq 0\right] \\
&= E\left[\frac{\sigma_t}{\sqrt{h_t}}E\left[|u_t|\right]|Z_{t-1} \neq 0\right] - E|u_t|E\left[\frac{\sigma_t}{\sqrt{h_t}}|Z_{t-1} \neq 0\right] \\
&= 0.
\end{aligned}$$

We used the independence between u_t and Z_{t-1} and the independence between u_t and $\frac{\sigma_t}{\sqrt{h_t}}$. In consequence, we have $\mathbb{P}[Z_{t-1} \neq 0] = 0$ and thus, almost surely $\lambda'\phi_t = 0$. We have

$$\frac{\partial \sigma_t^2}{\partial \theta} = \begin{pmatrix} 1 \\ \epsilon_{t-1}^2 \\ \sigma_{t-1}^2 \end{pmatrix} + \beta \frac{\partial \sigma_{t-1}^2}{\partial \theta}.$$

Using the stationarity of $\frac{\sigma_t^2}{\partial \theta}$ we obtain almost surely

$$\lambda' \begin{pmatrix} 1 \\ \epsilon_{t-1}^2 \\ \sigma_{t-1}^2 \end{pmatrix} = 0.$$

Since ϵ_{t-1}^2 cannot be measurable with respect to the σ-field generated by $\{u_k, k < t-1\}$, we have $\lambda_2 = 0$, then $\lambda_3 = 0$ and finally we have obtained $\lambda = 0_3$ and thus the inversibility of the matrix $B(\theta_0)$.

□

Lemma 4.6.7. *Under the assumptions of Theorem 4.3.2,*

$$\left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \frac{\partial l_t(\theta_0)}{\partial \theta} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} \right\} \right\| \to 0$$

$$\sup_{\theta \in V(\theta_0)} \left\| \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta \partial \theta'} \right\} \right\| \to 0,$$

where the convergence are in probability when $n \to +\infty$.

Proof. This Lemma can be proven in the exact same way as in the proof for the well specified GARCH model. □

With the previous lemmas, we are now able to prove Theorem 4.3.2. Using a Taylor expansion, we have for $(i, j) \in \{1, \cdots, 3\}^2$

$$0 = \frac{\partial}{\partial \theta} \tilde{I}_n(\theta_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} + \left\{ \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 \tilde{l}_t(\theta_i^*)}{\partial \theta_i \partial \theta_j} \right\} \sqrt{n}(\theta_n - \theta_0).$$

Using Lemmas 4.6.3 and 4.6.5, the Cramér-Wold theorem and the central limit theorem for square integrable stationary martingale difference of Billingsley (1995), we obtain

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \theta} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, B(\theta_0)\right).$$

Then using Lemma 4.6.5, Theorem 4.3.1 and the ergodic theorem we obtain

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t(\theta_{ij}^*)}{\partial \theta_i \partial \theta_j} \to A(\theta_0)(i, j), \quad a.s.$$

Finally, using Lemma 4.6.7, we conclude and obtain (4.3.5).

4.6.6 Proof of Theorem 4.4.1

From (4.4.3) and for $\theta \in \Theta$, we have

$$I_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left\{ \log \frac{\omega + \alpha \epsilon_{t-1}^2}{w(\Delta_t) + a(\Delta_t) \epsilon_{t-1}^2} + u_t^2 \frac{w(\Delta_t) - \omega + \left(a(\Delta_t) - \alpha\right) \epsilon_{t-1}^2}{\omega + \alpha \epsilon_{t-1}^2} \right\}.$$

From Remark 4.2.1 (which is also true in the ARCH(1) case) and the assumption that $\gamma > 0$, we have $h_t \to +\infty$ almost surely when $t \to +\infty$. Moreover, we have $\log \epsilon_t^2 = \log h_t + \log u_t^2$. The assumption $\gamma > 0$ implies that $E \log u_t^2 < +\infty$. In consequence, we have

$$\epsilon_t^2 \to +\infty, \ a.s. \text{ when } t \to +\infty. \tag{4.6.23}$$

Therefore, we have

$$I_n(\theta) = Q_n(\alpha) + K_t + R_{1n}(\theta) + R_{2n}(\theta),$$

with

$$\begin{aligned}
Q_n(\alpha) &= \log \alpha + \frac{1}{\alpha}\frac{1}{n}\sum_{t=1}^{n} u_t^2 a(\Delta_t) \\
R_{1n}(\theta) &= \frac{1}{n}\sum_{t=1}^{n}\left\{\log \frac{\omega + \alpha\epsilon_{t-1}^2}{w(\Delta_t) + a(\Delta_t)\epsilon_{t-1}^2} - \log \frac{\alpha}{a(\Delta_t)}\right\} \\
R_{2n}(\theta) &= \frac{1}{n}\sum_{t=1}^{n} u_t^2 \left\{\frac{w(\Delta_t) - \omega + (a(\Delta_t) - \alpha)\epsilon_{t-1}^2}{\omega + \alpha\epsilon_{t-1}^2} - \frac{a(\Delta_t) - \alpha}{\alpha}\right\}.
\end{aligned}$$

With (4.6.23), the Cesàro Lemma and the compactness of the parameter space, we obtain for $i \in \{1, 2\}$

$$\sup_{\theta\in\Theta} |R_{in}(\theta)| \to 0, \quad \text{a.s. when } n \to +\infty. \tag{4.6.24}$$

It is easy to obtain that the function Q_n admits a unique minimizer $\tilde{\alpha}_n$ characterized by

$$\tilde{\alpha}_n = \frac{1}{n}\sum_{t=1}^{n} u_t^2 a(\Delta_t). \tag{4.6.25}$$

We prove that almost surely

$$|\alpha_n - \tilde{\alpha}_n| \to 0, \quad \text{a.s.} \tag{4.6.26}$$

For any $n \in \mathbb{N}$, we define $\tilde{\theta}_n = (\tilde{\omega}_n, \tilde{\alpha}_n) \in \Theta$. Using the definitions of θ_n and $\tilde{\theta}_n$ and (4.6.24), we obtain

$$Q_n(\alpha_n) - Q_n(\tilde{\alpha}_n) \to 0, \quad \text{a.s. when } n \to 0. \tag{4.6.27}$$

Since Q_n is continuous and since $\tilde{\alpha}_n$ is its unique minimizer, we obtain $\alpha_n - \tilde{\alpha}_n \to 0$ almost surely. Hence we have (4.4.4) and the convergence (4.4.5) is obtained by the law of large numbers. Finally (4.4.6) is obtained by the central limit theorem.

4.6.7 Proof of Theorem 4.4.2

We define $\underline{\omega} = \inf \{\omega|\theta \in \Theta\}$ and $\overline{\omega} = \sup \{\omega|\theta \in \Theta\}$. The quantities $\underline{\alpha}$, $\overline{\alpha}$, $\underline{\beta}$ and $\overline{\beta}$ are defined in the same way.

Lemma 4.6.8. *We define $\Theta_\gamma = \{\theta \in \Theta, \ \beta < e^\gamma\}$. Under the assumptions of Theorem 4.4.2, for* *any* $\theta \in \Theta_\gamma$, $v_t(\alpha,\beta)$ *is stationary and ergodic. Moreover, for any compact* $\Theta_\gamma^* \subset \Theta_\gamma$

$$\sup_{\theta\in\Theta_\gamma^*}\left|\frac{\sigma_t^2(\theta)}{h_t} - v_t(\alpha,\beta)\right| \to 0 \ a.s.\ when\ t\to+\infty. \tag{4.6.28}$$

If $\theta \notin \Theta_\gamma$, *we have*

$$\frac{\sigma_t^2}{h_t} \to +\infty,\ \ a.s.\ \ when\ t\to+\infty. \tag{4.6.29}$$

Proof. We begin by proving the stationarity and ergodicity of v_t. Writing the Cauchy root test for the series $v_t(\alpha,\beta)$, we obtain

$$\left\{\alpha\beta^{j-1}u_{t-j}^2\prod_{k=1}^{j}\frac{1}{a_{t-k}}\right\}^{1/j} = \exp\left\{\frac{1}{j}\log(\alpha u_{t-j}^2) + \frac{j-1}{j}\log\beta - \frac{1}{j}\sum_{k=1}^{j}a_{t-k}\right\}.$$

Using the stationarity of the process (u_t) and the ergodicity of the process (a_t), we obtain that $(v_t(\alpha,\beta))$ is a.s. finite if

$$\log\beta+\gamma<0 \ \Leftrightarrow\ \beta<e^\gamma.$$

Therefore for $\theta\in\Theta_\gamma$, the series $v_t(\alpha,\beta)$ is a.s. finite and $v_t(\alpha,\beta)$ is thus a measurable function of $\{u_{t'},\ t'<t\}$ and $v_t(\alpha,\beta)$ is stationary and ergodic.

We now prove (4.6.28), we have

$$\frac{\sigma_t^2}{h_t} = \sum_{j=1}^{t}\beta^{j-1}\left\{\prod_{k=1}^{j}\frac{h_{t-k}}{h_{t-k+1}}\right\}\frac{\omega+\alpha\epsilon_{t-j}^2}{h_{t-j}} = q_{1t}+q_{2t}, \tag{4.6.30}$$

where

$$q_{1t} = \sum_{j=1}^{t}\beta^{j-1}\frac{\omega}{h_t},\ \text{and}\ q_{2t} = \sum_{j=1}^{t}\beta^{j-1}\alpha u_{t-j}^2\left\{\prod_{k=1}^{j}\frac{h_{t-k}}{h_{t-k+1}}\right\}.$$

We define $\overline{\beta}_\gamma = \sup\{\beta|\theta\in\Theta_\gamma^*\}$ ($\overline{\alpha}_\gamma$ is defined in the same way), we have

$$q_{1t} \le K\overline{w}\frac{t+\overline{\beta}_\gamma^t}{h_t}.$$

Thus, by Proposition 4.4.1 and Assumption **B0**, we obtain $\sup_{\theta\in\Theta_\gamma} q_{1t}\to 0$, almost surely when

$t \to +\infty$. We have,

$$\frac{h_{t-k}}{h_{t-k+1}} - \frac{1}{a_{t-k}} = -\frac{w(\Delta_{t-k+1})}{a_{t-k}h_{t-k+1}}.$$

In consequence, we obtain $\frac{h_{t-k}}{h_{t-k+1}} = \frac{1}{a_{t-k}}\left(1 - \frac{w(\Delta_{t-k+1})}{h_{t-k+1}}\right)$ and

$$\prod_{k=1}^{j}\frac{1}{a_{t-k}} - \prod_{k=1}^{j}\frac{h_{t-k}}{h_{t-k+1}} = \left(\prod_{k=1}^{j}\frac{1}{a_{t-k}}\right)\left\{1 - \prod_{k=1}^{j}\left(1 - \frac{w(\Delta_{t-k+1})}{h_{t-k+1}}\right)\right\}. \tag{4.6.31}$$

We have

$$0 \le v_t - q_{2t} \le \sum_{j=1}^{t_0}\alpha u_{t-j}^2\beta^{j-1}\left\{\prod_{k=1}^{j}\frac{1}{a_{t-k}} - \prod_{k=1}^{j}\frac{h_{t-k}}{h_{t-k+1}}\right\} + \sum_{j=t_0+1}^{+\infty}\alpha u_{t-j}^2\beta^{j-1}\prod_{k=1}^{j}\frac{1}{a_{t-k}}.$$

The series $\sum_j \overline{\alpha}_\gamma u_{t-j}^2\overline{\beta}_\gamma^{j-1}\prod_{k=1}^{j}\frac{1}{a_{t-k}}$ is convergent, thus $\sum_{j=t_0}^{+\infty}\overline{\alpha}_\gamma u_{t-j}^2\overline{\beta}_\gamma^{j-1}\prod_{k=1}^{j}\frac{1}{a_{t-k}}$ converges almost surely toward zero when $t_0 \to +\infty$. With (4.6.31), we obtain that for any $t_0 \in \mathbb{N}$,

$$\sum_{j=1}^{t_0}\overline{\alpha}_\gamma u_{t-j}^2\overline{\beta}_\gamma^{j-1}\left\{\prod_{k=1}^{j}\frac{1}{a_{t-k}} - \prod_{k=1}^{j}\frac{h_{t-k}}{h_{t-k+1}}\right\} \to 0, \text{ a.s. when } t \to +\infty.$$

Finally, we have obtained

$$\sup_{\theta\in\Theta_\gamma}(v_t - q_{2t}) \to 0, \ a.s.$$

This yields (4.6.28).

Now, for $\theta \notin \Theta_\Gamma$, for any $t_0 < t$ and for $\rho > e^{-\gamma}$, we have

$$\frac{\sigma_t^2}{h_t} \ge \sum_{j=1}^{t_0}\alpha u_{t-j}^2\beta^{j-1}\prod_{k=1}^{j}\frac{1}{a_{t-k}} + o(\rho^t t_0^2) \text{ a.s.}$$

When $\beta > e^{-\gamma}$, this series is divergent by the Cauchy root test and when $\beta = e^{-\gamma}$, we also obtain the divergence of the series by the Chung-Fuchs Theorem and (4.6.29) is obtained. □

Lemma 4.6.9. *Under the assumptions of Theorem 4.4.2, we have*

$$\sup_{\theta\in\Theta}\frac{h_t}{\sigma_t^2} \le V_t,$$

with

$$V_t = \sum_{i=1}^{+\infty} \mathbb{1}_{|u_{t-1}|\leq\varepsilon} \cdots \mathbb{1}_{|u_{t-i+1}|\leq\varepsilon} \mathbb{1}_{|u_{t-i}|>\varepsilon} \left\{ \frac{\overline{w}}{\underline{\omega}} \sum_{j=0}^{i-2} \left(\frac{\overline{a}\varepsilon^2 + \overline{b}}{\underline{\beta}} \right)^j + \left(\frac{\overline{a}\varepsilon^2 + \overline{b}}{\underline{\beta}} \right)^{i-1} H(\varepsilon) \right\}.$$

Moreover, for any $y > 0$, we have

$$EV_t^y < +\infty \text{ and consequently } E\left(\sup_{\theta\in\Theta} \frac{h_t}{\sigma_t^2}\right)^y < +\infty. \tag{4.6.32}$$

Proof. For $\varepsilon > 0$, if $|u_{t-1}| > \varepsilon$, we have

$$\begin{aligned} \frac{h_t}{\sigma_t^2} &= \frac{w(\Delta_t) + (a(\Delta_t)u_{t-1}^2 + b(\Delta_t))h_{t-1}}{\omega + ah_{t-1}u_{t-1}^2 + b\sigma_{t-1}^2} \\ &\leq \frac{\overline{w}}{\underline{\omega}} + \frac{\overline{a}}{\underline{\alpha}} + \frac{\overline{b}}{\underline{\alpha}\varepsilon} = H(\varepsilon). \end{aligned} \tag{4.6.33}$$

Now, if $|u_{t-1}| \leq \varepsilon$ and if $|u_{t-2}| > \varepsilon$, from (4.6.33), we have

$$\begin{aligned} \frac{h_t}{\sigma_t^2} &\leq \frac{\overline{w}}{\underline{\omega}} + \frac{\overline{a}\varepsilon^2 + \overline{b}}{\underline{\beta}} \frac{h_{t-1}}{\sigma_{t-1}^2} \\ &\leq \frac{\overline{w}}{\underline{\omega}} + \frac{\overline{a}\varepsilon^2 + \overline{b}}{\underline{\beta}} H(\varepsilon). \end{aligned}$$

Iterating this mechanism, we obtain

$$\sup_{\theta\in\theta} \frac{h_t}{\sigma_t^2} \leq V_t.$$

We have $E\left[\mathbb{1}_{|u_{t-1}|\leq\varepsilon} \cdots \mathbb{1}_{|u_{t-i+1}|\leq\varepsilon} \mathbb{1}_{|u_{t-i}|>\varepsilon}\right] = (1 - p(\varepsilon))p(\varepsilon)^{i-1}$, where $p(\varepsilon) = \mathbb{P}\left[|u_t| \leq \varepsilon\right]$. Now, we have

$$\frac{\overline{w}}{\underline{\omega}} \sum_{j=0}^{i-2} \left(\frac{\overline{a}\varepsilon^2 + \overline{b}}{\underline{\beta}} \right)^j + \left(\frac{\overline{a}\varepsilon^2 + \overline{b}}{\underline{\beta}} \right)^{i-1} H(\varepsilon) \leq KH(\varepsilon) \left(\frac{\overline{a}\varepsilon^2 + \overline{b}}{\underline{\beta}} \right)^{i-1}.$$

Assumption **B3** yields $\lim_{\varepsilon\to 0} p(\varepsilon) = 0$, therefore for ε small enough, we have $p(\varepsilon)\frac{\overline{a}\varepsilon^2+\overline{b}}{\underline{\beta}} < 1$ and we obtain (4.6.32). The cases where $y > 1$ can be treated with the exact same reasoning. □

Lemma 4.6.10. *Under the assumptions of Theorem 4.4.2, we have*

$$E\left[\sup_{\theta\in\Theta} \left(\frac{1}{v_t(\alpha, \beta)}\right)^y\right] < +\infty. \tag{4.6.34}$$

Proof. The proof of this lemma is very similar to the proof of Lemma 4.6.9. For $\varepsilon > 0$, if $|u_{t-1}| > \varepsilon$ and since $v_t(\alpha,\beta)$ is greater than its first term, we have

$$\frac{1}{v_t(\alpha,\beta)} \leq \frac{a_{t-1}}{\alpha u_{t-1}^2} \leq \frac{\overline{a}}{\underline{\alpha}} + \frac{\overline{b}}{\underline{\alpha}\varepsilon} = G(\varepsilon).$$

Now, if $|u_{t-1}| \leq \varepsilon$ and $|u_{t-2}| > \varepsilon$, $v_t(\alpha,\beta)$ is greater than its second term and we obtain

$$\frac{1}{v_t(\alpha,\beta)} \leq \frac{a_{t-1}}{\beta}\frac{a_{t-2}}{\alpha u_{t-2}^2} \leq \frac{\overline{a}\varepsilon + \overline{b}}{\underline{\beta}} G(\varepsilon).$$

Iterating this reasoning, we obtain

$$\sup_{\theta\in\Theta} \frac{1}{v_t(\alpha,\beta)} \leq H(\varepsilon)\sum_{i=1}^{+\infty} \mathbb{1}_{|u_{t-1}|\leq\varepsilon}\cdots\mathbb{1}_{|u_{t-i+1}|\leq\varepsilon}\mathbb{1}_{|u_{t-i}|>\varepsilon}\left(\frac{\overline{a}\varepsilon+\overline{b}}{\underline{\beta}}\right)^{i-1}.$$

Consequently, we have, for any $y > 0$

$$E\sup_{\theta\in\Theta}\left(\frac{1}{v_t(\alpha,\beta)}\right)^y \leq G(\varepsilon)^y(1-p(\varepsilon))\sum_{i=1}^{+\infty} p(\varepsilon)^{i-1}\left(\frac{\overline{a}\varepsilon+\overline{b}}{\underline{\beta}}\right)^{y(i-1)}.$$

And we conclude as in the proof of Lemma 4.6.9. □

In order to prove Theorem 4.4.2, we first prove that, for any compact $\Theta_\gamma^* \subset \Theta_\gamma$

$$\lim_{n\to+\infty}\sup_{\theta\in\Theta_\gamma^*} |R_n(\theta)| = 0, \text{ a.s.} \tag{4.6.35}$$

We have

$$\begin{aligned} R_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n}\left\{u_t^2\left(\frac{h_t}{\sigma_t^2} - \frac{1}{v_t(\alpha,\beta)}\right) + \log\frac{\sigma_t^2}{h_t v_t(\alpha,\beta)}\right\} \\ &= \frac{1}{n}\sum_{t=1}^{n}\left\{u_t^2\frac{h_t}{\sigma_t^2}\frac{1}{v_t(\alpha,\beta)}\left(v_t(\alpha,\beta) - \frac{\sigma_t^2}{h_t}\right) + \log\frac{\sigma_t^2}{h_t} - \log v_t(\alpha,\beta)\right\}. \end{aligned}$$

Using Lemma 4.6.9, we obtain that for any $\theta \in \Theta_\gamma^*$

$$|R_n(\theta)| \leq \frac{1}{n}\sum_{t=1}^{n}\left\{u_t^2\frac{V_t}{v_t(\underline{\alpha},\underline{\beta})}\left|v_t - \frac{\sigma_t^2}{h_t}\right| + K\left|v_t - \frac{\sigma_t^2}{h_t}\right|\right\}.$$

Lemma 4.6.8 yields, for any $\varepsilon > 0$ the existence of t_0 such that, for any $t \geq t_0$, we have almost surely $\left| v_t - \frac{\sigma_t^2}{h_t} \right| < \varepsilon$. Therefore, we have, for $n \geq t_0$

$$|R_n(\theta)| \leq \frac{1}{n} \sum_{t=1}^{t_0-1} \left\{ \left(u_t^2 \frac{V_t}{v_t(\underline{\alpha}, \underline{\beta})} + K \right) \left| v_t - \frac{\sigma_t^2}{h_t} \right| \right\} + \varepsilon \left(K + \frac{1}{n} \sum_{t=t_0}^{n} u_t^2 \frac{V_t}{v_t(\underline{\alpha}, \underline{\beta})} \right).$$

Using Lemma 4.6.9, Lemma 4.6.10 and the ergodic theorem, we conclude and obtain (4.6.35). The case where $\theta \notin \Theta_\gamma$ is easily treated. In this case, with (4.6.29) we have $I_n(\theta) \to +\infty$.

With the stationarity and the ergodicity of the process $(v_t(\alpha, \beta))$ (for $\theta \in \Theta_\gamma$), we obtain

$$\lim_{n \to +\infty} O_n(\alpha, \beta) = E \left[u_t^2 \left(\frac{1}{v_t(\alpha, \beta)} - 1 \right) + \log v_t(\alpha, \beta) \right], \text{ a.s.}$$

With Assumption **B1** and standard arguments using the compactness of the parameter space Θ we conclude and obtain (4.4.12).

4.6.8 Proof of Remark 4.4.1

We define $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ as

$$\left(\tilde{\alpha}_n, \tilde{\beta}_n \right) = \underset{(\alpha,\beta) \in \Theta_{\alpha,\beta}}{\operatorname{argmin}} O_n(\alpha, \beta),$$

where $\Theta_{\alpha\beta} = \left\{ (\alpha, \beta) |\ \exists\ \omega > 0,\ (\omega, \alpha, \beta)' \in \Theta \right\}$. We remark that for any $(\alpha, \beta) \in \Theta_{\alpha,\beta}$, $v_t(\alpha, \beta)$ can be decomposed as $v_t(\alpha, \beta) = \alpha F_t(\beta)$ where $F_t(\beta) = \sum_{j=1}^{+\infty} \beta^{j-1} u_{t-j}^2 \prod_{k=1}^{j-1} \frac{1}{a_{t-k}}$. For any fixed value of β, we have

$$\begin{aligned} \underset{\alpha}{\operatorname{argmin}}\ O_n(\alpha, \beta) &= \underset{\alpha}{\operatorname{argmin}}\ \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{1}{\alpha} \frac{u_t^2}{F_t(\beta)} + \log \alpha \right\} \\ &= \frac{1}{n} \sum_{t=1}^{n} \frac{u_t^2}{F_t(\beta)}. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} \inf_{(\alpha,\beta) \in \Theta_{\alpha,\beta}} O_n(\alpha, \beta) &= \inf_{\beta} O_n \left(\frac{1}{n} \sum_{t=1}^{n} \frac{u_t^2}{F_t(\beta)}, \beta \right) \\ &= \inf_{\beta} \left\{ \log \left(\frac{1}{n} \sum_{t=1}^{n} \frac{u_t^2}{F_t(\beta)} \right) + \frac{1}{n} \sum_{t=1}^{n} \log F_t(\beta) \right\}. \end{aligned}$$

Thus, we have $\tilde{\beta}_n = \text{argmin} \left\{\log\left(\frac{1}{n}\sum_{t=1}^n \frac{u_t^2}{F_t(\beta)}\right) + \frac{1}{n}\sum_{t=1}^n \log F_t(\beta)\right\}$ and $\tilde{\alpha}_n = \frac{1}{n}\sum_{t=1}^n \frac{u_t^2}{F_t(\tilde{\beta}_n)}$. Using the ergocitity of $(F_t(\beta))_t$ (proven in Lemma 4.6.8, if $\beta < e^{\gamma}$), we obtain, when $n \to +\infty$,

$$\log\left(\frac{1}{n}\sum_{t=1}^n \frac{u_t^2}{F_t(\beta)}\right) + \frac{1}{n}\sum_{t=1}^n \log F_t(\beta) \to \log E\left[\frac{1}{F_t(\beta)}\right] - E\left[\log \frac{1}{F_t(\beta)}\right], \quad \text{a.s.}$$

The last convergence cannot be directly proven and requires the different lemmas given in Subsection 4.6.7. Assumption **B1'** and standard arguments yields $\tilde{\beta}_n \to \beta_0$ almost surely and consequently we have $\tilde{\beta}_n \to \beta_0$. Defining $\alpha_0 = E\left[\frac{1}{F_t(\beta_0)}\right]$, we also obtain the almost sure convergence $\tilde{\alpha}_n \to \alpha_0$. Then, with similar arguments as those used with Assumption **B1**, we conclude and obtain (4.4.12).

4.6.9 Proof of Theorem 4.4.3

Lemma 4.6.11. *For any $y > 0$ and for any $\theta = (\omega, \alpha, \beta) \in \Theta^{(y)}$, we have*

$$\|v_t(\alpha,\beta)\|_y < +\infty, \text{ and } \left\|v_t(\alpha,\beta) - \frac{\sigma_t^2(\theta)}{h_t}\right\|_y \to 0, \text{ as } t \to +\infty. \tag{4.6.36}$$

Proof. Let $y > 0$ and let $\theta \in \Theta^{(y)}$, the first part in (4.6.36) is true if the series $\sum_j \left\|\prod_{k=1}^{j-1} \frac{\beta}{a_{t-k}}\right\|_y$ is convergent. With already used arguments, we obtain

$$\begin{aligned}
E\left[\prod_{k=1}^{j-1}\left(\frac{\beta}{a_{t-k}}\right)^y\right] &= \beta^{j-1} E\left[E\left[\prod_{k=1}^{j-1}\frac{1}{a_{t-k}^y}|\Delta_t, \cdots, \Delta_{t-j+1}\right]\right] \\
&= \beta^{j-1} E\left[\prod_{k=1}^{j-1} \tilde{a}^{(y)}(\Delta_{t-k+1})\right] \\
&= \beta^{j-1} \mathbb{P}_{\tilde{a}^{(y)}}^{j-1} \Pi_1.
\end{aligned}$$

We have $\theta \in \Theta^{(y)}$, therefore using the Jordan decomposition and the definition of $\Theta^{(y)}$, we obtain that the series is convergent and thus the first part of (4.6.36). Replacing the almost sure convergences by convergences in L_y, the second part of (4.6.36) can be obtained with the same arguments as those used to obtain the almost sure convergence of $|\frac{\sigma_t^2}{h_t} - v_t|$ in Lemma 4.6.8. $\square$

Lemma 4.6.12. *Let ϖ be an arbitrary compact subset of $(0, +\infty)$. Under the assumptions of*

Theorem 4.4.3, we have for any $y \in \mathbb{N}$

$$\sup_{\omega\in\varpi} \sum_{t=1}^{+\infty} \left\| v_t(\alpha_0,\beta_0) - \frac{\sigma_t^2(\omega,\alpha_0,\beta_0)}{h_t} \right\|_y < +\infty. \tag{4.6.37}$$

Proof. For simplicity purpose, we will only prove the result on the case $y = 1$, the other cases can be obtained with the same reasoning. In all the proof of this lemma, the quantity are taken at $\theta = (\omega, \alpha_0, \beta_0)$. From (4.6.30), we have

$$\left| v_t - \frac{\sigma_t^2}{h_t} \right| \leq q_{1t} + \left| v_t - q_{2t} \right|.$$

We treat the second term of last equation, we have

$$0 \leq v_t - q_{2t} \leq \sum_{j=1}^{\lfloor t/2 \rfloor} \alpha\beta^{j-1} u_{t-j}^2 \left\{ \prod_{k=1}^{j} \frac{1}{a_{t-k}} - \prod_{k=1}^{j} \frac{h_{t-k}}{h_{t-k+1}} \right\} + \sum_{j=\lfloor t/2 \rfloor+1}^{+\infty} \alpha\beta^{j-1} u_{t-j}^2 \prod_{k=1}^{j} \frac{1}{a_{t-k}}. \tag{4.6.38}$$

From (4.6.31), we have

$$\alpha\beta^{j-1} u_{t-j}^2 \left\{ \prod_{k=1}^{j} \frac{1}{a_{t-k}} - \prod_{k=1}^{j} \frac{h_{t-k}}{h_{t-k+1}} \right\} \leq K \left\{ 1 - \prod_{k=1}^{j} \left(1 - \frac{\overline{w}}{h_{t-k+1}} \right) \right\}. \tag{4.6.39}$$

For $j \leq \lfloor t/2 \rfloor$, we have

$$1 - \prod_{k=1}^{j} \left(1 - \frac{\overline{w}}{h_{t-k+1}} \right) \leq 1 - \prod_{k=1}^{\lfloor t/2 \rfloor} \left(1 - \frac{\overline{w}}{h_{t-k+1}} \right). \tag{4.6.40}$$

If $h_0 = 0$, we have $h_t = \sum_{k=1}^{t} a_t \cdots a_{t-k+1} w(\Delta_{t-k+1})$. Therefore we have

$$\begin{aligned} \epsilon_{t+1}^2 &= u_{t+1}^2 \sum_{k=1}^{t+1} a_{t+1} \cdots a_{t-k+2} w(\Delta_{t-k+2}) \\ &\geq u_{t+1}^2 \sum_{k=1}^{t} a_{t+1} \cdots a_{t-k+2} w(\Delta_{t-k+2}) \overset{d}{=} \epsilon_t^2. \end{aligned}$$

Thus, we obtain

$$E\left[\frac{1}{h_{t-1}}\frac{1}{h_t}\right] - E\left[\frac{1}{h_{t-1}^2}\right] \leq E\left[\frac{1}{\underline{w}}\left(\frac{1}{h_t} - \frac{1}{h_{t-1}}\right)\right] \leq 0.$$

In consequence, we have

$$\begin{aligned} E\left[1 - \prod_{k=1}^{\lfloor t/2\rfloor}\left(1 - \frac{\overline{w}}{h_{t-k+1}}\right)\right] &\leq E\left[\left(1 + \frac{\overline{w}}{h_{\lfloor t/2\rfloor}}\right)^{\lfloor t/2\rfloor} - 1\right] \\ &\leq \sum_{k=1}^{\lfloor t/2\rfloor}\binom{\lfloor t/2\rfloor}{k} E\left[\frac{\overline{w}^k}{h_{\lfloor t/2\rfloor}^k}\right]. \end{aligned} \tag{4.6.41}$$

For any $t > 0$, we have $h_t \geq a_t \cdots a_1 w(\Delta_1)$, thus with already used arguments we obtain that there exists $\rho_k \in (\rho(\mathbb{P}_{\tilde{a}^{(k)}}), 1)$ such that

$$E\left[\frac{1}{h_t^k}\right] \leq K\rho_k^t.$$

With Assumption **B5** and choosing ρ arbitrarily close to $\rho(\mathbb{P}_{\tilde{a}^{(k)}})$ we obtain the existence of ρ such that

$$E\left[\frac{1}{h_t^k}\right] < K\rho^{tk}. \tag{4.6.42}$$

With (4.6.41) and (4.6.42), we have

$$E\left[1 - \prod_{k=1}^{\lfloor t/2\rfloor}\left(1 - \frac{\overline{w}}{h_{t-k+1}}\right)\right] \leq K\sum_{k=1}^{\lfloor t/2\rfloor}\binom{\lfloor t/2\rfloor}{k}\overline{w}^k\rho^{k\lfloor t/2\rfloor}. \tag{4.6.43}$$

Each term of the last sum can be bounded as follows

$$\binom{\lfloor t/2\rfloor}{k}\overline{w}^k\rho^{k\lfloor t/2\rfloor} \leq \lfloor t/2\rfloor^k\overline{w}^k\rho^{\lfloor t/2\rfloor k} = \exp\left\{k\left(\log\lfloor t/2\rfloor + \overline{w} + \lfloor t/2\rfloor\log\rho\right)\right\}.$$

Define t_0 such that for any $t \geq t_0$, we have $\log\lfloor t/2\rfloor + \overline{w} + \lfloor t/2\rfloor\log\rho < 0$. For $t \geq t_0$, we have

$$\sum_{k=1}^{\lfloor t/2\rfloor}\binom{\lfloor t/2\rfloor}{k}\overline{w}^k\rho^{k\lfloor t/2\rfloor} \leq \lfloor t/2\rfloor\exp\left\{\log\lfloor t/2\rfloor + \overline{w} + \lfloor t/2\rfloor\log\rho\right\}.$$

With (4.6.39), (4.6.40), (4.6.41) and (4.6.43) we have obtained

$$\begin{aligned} E\left[\sum_{j=1}^{\lfloor t/2\rfloor}\alpha\beta^{j-1}u_{t-j}^2\left\{\prod_{k=1}^{j}\frac{1}{a_{t-k}}-\prod_{k=1}^{j}\frac{h_{t-k}}{h_{t-k+1}}\right\}\right] &\leq K\sum_{j=1}^{\lfloor t/2\rfloor}E\left[1-\prod_{k=1}^{j}\left(1-\frac{\overline{w}}{h_{t-k+1}}\right)\right] \\ &\leq \lfloor t/2\rfloor E\left[1-\prod_{k=1}^{\lfloor t/2\rfloor}\left(1-\frac{\overline{w}}{h_{t-k+1}}\right)\right] \\ &\leq \lfloor t/2\rfloor^2\exp\left\{\log\lfloor t/2\rfloor+\overline{w}+\lfloor t/2\rfloor\log\rho\right\}. \end{aligned}$$

Finally, we have proved

$$\sum_{t=1}^{+\infty}E\left[\sum_{j=1}^{\lfloor t/2\rfloor}\alpha\beta^{j-1}u_{t-j}^2\left\{\prod_{k=1}^{j}\frac{1}{a_{t-k}}-\prod_{k=1}^{j}\frac{h_{t-k}}{h_{t-k+1}}\right\}\right]<+\infty. \tag{4.6.44}$$

We now have to treat the second term in (4.6.38). We have

$$E\left[\alpha\beta^{j-1}u_{t-j}^2\prod_{k=1}^{j}\frac{1}{a_{t-k}}\right]\leq K\beta^{j-1}E\left[\prod_{k=1}^{j-1}\frac{1}{a_{t-k}}\right].$$

From Assumption **B4**, we obtain $E\left[\alpha\beta^{j-1}u_{t-j}^2\prod_{k=1}^{j}\frac{1}{a_{t-k}}\right]\leq K\rho^{j-1}$ and in consequence

$$E\left[\sum_{j=\lfloor t/2\rfloor+1}^{+\infty}\alpha\beta^{j-1}u_{t-j}^2\prod_{k=1}^{j}\frac{1}{a_{t-k}}\right]<K\rho^{\lfloor t/2\rfloor}.$$

Using (4.6.44), last equation and similar arguments to treat $\sum Eq_{1t}$, we obtain (4.6.37). $\square$

Lemma 4.6.13. *Under the assumptions of Theorem 4.4.3, for $i\in\{2,3\}$ and for $\theta\in\Theta_\gamma$, the process $\left(\frac{\partial v_t}{\partial\theta_i}\right)_t$ is stationary, ergodic and satisfies*

$$\left|\frac{\partial v_t}{\partial\theta_i}-\frac{1}{h_t}\frac{\partial\sigma_t^2(\theta)}{\partial\theta_i}\right|\to 0,\ \ a.s. \tag{4.6.45}$$

Moreover for any $y>0$ and for any $\theta\in\Theta^{(y)}$, we have

$$\left\|\frac{\partial v_t}{\partial\theta_i}\right\|_y<+\infty,\ \ and\ \ \left\|\frac{\partial v_t}{\partial\theta_i}-\frac{1}{h_t}\frac{\partial\sigma_t^2(\theta)}{\partial\theta_i}\right\|_y\to 0,\ \ as\ t\to+\infty, \tag{4.6.46}$$

and

$$\sum_{t=1}^{+\infty} \left\| \frac{\partial v_t}{\partial \theta_i} - \frac{1}{h_t} \frac{\partial \sigma_t^2(\theta)}{\partial \theta_i} \right\|_y < +\infty. \tag{4.6.47}$$

Proof. The ergodicity, the stationarity of $\left(\frac{\partial v_t}{\partial \theta_i}\right)_t$ and the almost sure convergence (4.6.45) can be obtained with the same arguments as in the proof of Lemma 4.6.8. Equation (4.6.46) can be obtained with the same arguments as in the proof of Lemma 4.6.11. Finally, (4.6.47) can be obtained with the same arguments as in the proof of Lemma 4.6.12. □

Lemma 4.6.14. *Let ϖ be an arbitrary compact subset of* $(0, +\infty)$. *Under the assumptions of Theorem 4.4.3, we have for $i \in \{2, 3\}$*

$$\sup_{\omega \in \varpi} \sum_{t=1}^{+\infty} E\left[\frac{\partial}{\partial \theta_i} l_t(\omega, \alpha_0, \beta_0)\right] < +\infty. \tag{4.6.48}$$

Proof. In this proof, all the quantities are taken at $\theta = (\omega, \alpha_0, \beta_0)$. From (4.4.10) and (4.4.11), we have $l_t = o_t + r_t$. We have

$$\begin{aligned} \frac{\partial o_t}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\left(u_t^2\left(\frac{1}{v_t} - 1\right) + \log v_t\right) \\ &= \frac{1}{v_t}\frac{\partial v_t}{\partial \alpha} - u_t^2 \frac{1}{v_t^2}\frac{\partial v_t}{\partial \alpha} = \frac{1}{\alpha}\left(1 - \frac{u_t^2}{v_t}\right). \end{aligned}$$

Using Assumption **B1** and similar arguments as in the proof of Lemma 4.6.3, we obtain for $i \in \{2, 3\}$

$$E\left[\frac{\partial o_t}{\partial \theta_i}\right] = E\left[\left(1 - \frac{u_t^2}{v_t}\right)\frac{1}{v_t}\frac{\partial v_t}{\partial \theta_i}\right] = \frac{\partial}{\partial \theta_i} E\left[o_t\right] = 0. \tag{4.6.49}$$

Besides, we have for $i \in \{2, 3\}$

$$\begin{aligned} \frac{\partial}{\partial \theta_i}\left(\log \frac{\sigma_t^2}{h_t v_t}\right) &= \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta_i} - \frac{1}{v_t}\frac{\partial v_t}{\partial \theta_i} \\ &= \frac{h_t}{\sigma_t^2}\left(\frac{1}{h_t}\frac{\partial \sigma_t^2}{\partial \theta_i} - \frac{\partial v_t}{\partial \theta_i}\right) + \frac{h_t}{\sigma_t^2}\frac{1}{v_t}\frac{\partial v_t}{\partial \theta_i}\left(v_t - \frac{\sigma_t^2}{h_t}\right), \end{aligned} \tag{4.6.50}$$

and

$$\frac{\partial}{\partial \theta_i}\left(u_t^2\left(\frac{h_t}{\sigma_t^2} - \frac{1}{v_t}\right)\right) = u_t^2\left\{\frac{h_t^2}{\sigma_t^4}\left(\frac{1}{h_t}\frac{\partial \sigma_t^2}{\partial \theta_i} - \frac{\partial v_t}{\partial \theta_i}\right) + \frac{\partial v_t}{\partial \theta_i}\frac{h_t}{\sigma_t^2}\frac{1}{v_t}\left(\frac{h_t}{\sigma_t^2} + \frac{1}{v_t}\right)\left(v_t - \frac{\sigma_t^2}{h_t}\right)\right\}. \tag{4.6.51}$$

Therefore using (4.6.49), (4.6.50), (4.6.51), Lemma 4.6.9, Lemma 4.6.10, Lemma 4.6.12, Lemma 4.6.13 and the Cauchy-Schwartz inequality, we obtain with already used arguments (4.6.48).

□

Lemma 4.6.15. *Under the assumptions of Theorem 4.4.3, for any $\omega > 0$, we have*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \frac{\partial}{\partial\alpha} l_t(\omega, \alpha_0, \beta_0) \\ \frac{\partial}{\partial\beta} l_t(\omega, \alpha_0, \beta_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, A_{\alpha_0,\beta_0}\right), \tag{4.6.52}$$

Moreover, the matrix A_{α_0,β_0} is definite positive.

Proof. We have $\frac{\partial}{\partial\theta} l_t = \left(1 - u_t^2 \frac{h_t}{\sigma_t^2}\right) \frac{h_t}{\sigma_t^2} \frac{1}{h_t} \frac{\partial \sigma_t^2}{\partial\theta}$. For $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$and for $\omega > 0$, we define $\nabla_t = \left(\frac{\partial}{\partial\alpha} l_t(\omega, \alpha_0, \beta_0), \frac{\partial}{\partial\beta} l_t(\omega, \alpha_0, \beta_0)\right) \lambda$ and we have

$$\begin{aligned} \nabla_t &= \left(1 - \frac{u_t^2}{v_t}\right) \frac{1}{v_t} \left(\lambda_1 \frac{\partial v_t}{\partial\alpha} + \lambda_2 \frac{\partial v_t}{\partial\beta}\right) \\ &+ \left(1 - u_t^2 \frac{h_t}{\sigma_t^2}\right) \frac{h_t}{\sigma_t^2} \left\{\lambda_1 \left(\frac{1}{h_t} \frac{\partial \sigma_t^2}{\partial\alpha} - \frac{\partial v_t}{\partial\alpha}\right) + \lambda_2 \left(\frac{1}{h_t} \frac{\partial \sigma_t^2}{\partial\beta} - \frac{\partial v_t}{\partial\beta}\right)\right\} \\ &+ \left(\lambda_1 \frac{\partial v_t}{\partial\alpha} + \lambda_2 \frac{\partial v_t}{\partial\beta}\right) \frac{h_t}{\sigma_t^2} \frac{1}{v_t} \left\{\frac{h_t}{\sigma_t^2} u_t^2 + 1 - \frac{u_t^2}{v_t}\right\} \left(\frac{\sigma_t^2}{h_t} - v_t\right). \end{aligned}$$

We define $\eta_{nt} = \frac{1}{\sqrt{n}} \left(\nabla_t - E\left[\nabla_t\right]\right)$. With Assumption **B4**, Lemma 4.6.11, Lemma 4.6.13 and noting that $\left|1 - u_t^2 \frac{h_t}{\sigma_t^2}\right| \leq \left|1 + u_t^2 V_t\right|$, we obtain

$$E\eta_{nt}^2 = \frac{1}{n} E\left[\left(1 - \frac{u_t^2}{v_t}\right)^2 \frac{1}{v_t^2} \left(\lambda_1 \frac{\partial v_t}{\partial\alpha} + \lambda_2 \frac{\partial v_t}{\partial\beta}\right)^2\right] + o(1), \text{ when } n \to +\infty.$$

Therefore, using the stationarity of (v_t), (d_t^α) and (d_t^β), we have

$$\sum_{t=1}^{n} E\eta_{nt}^2 \to E\left[\left(1 - \frac{u_1^2}{v_1}\right)^2 \frac{1}{v_1^2} \left(\lambda_1 \frac{\partial v_t}{\partial\alpha} + \lambda_2 \frac{\partial v_t}{\partial\beta}\right)^2\right] < +\infty, \text{ when } n \to +\infty$$

In order to use the central limit theorem for martingales differences as in the proof of Theorem 4.3.2, it remains to prove the Lindeberg condition $\sum_{t=1}^{n} E\left[\eta_{nt}^2 \mathbb{1}_{|\eta_{nt}|>\varepsilon}\right] \to 0$ when $n \to +\infty$.

Using the Hölder inequality and the Markov inequality, we have

$$\begin{aligned}
E\left[\eta_{nt}^2 \mathbb{1}_{|\eta_{nt}|>\varepsilon}\right] &\leq \left\|\eta_{nt}^2\right\|_{1+\eta_1} E\left[\mathbb{1}_{|\eta_{nt}|>\varepsilon}\right]^{\frac{\eta_1}{1+\eta_1}} \\
&\leq \left\|\eta_{nt}^2\right\|_{1+\eta_1} \frac{1}{\varepsilon} E\left[|\eta_{nt}|\right]^{\frac{\eta_1}{1+\eta_1}} \\
&\leq K \frac{1}{n\sqrt{n}} \left\{\|\nabla_t^2\|_{1+\eta_1} + E\left[\nabla_t\right]^2\right\} E\left[|\nabla_t|\right]^{\frac{\eta_1}{1+\eta_1}}.
\end{aligned}$$

With (4.6.36) and (4.6.46), we obtain

$$\begin{aligned}
\left\|\nabla_t^2\right\|_{1+\eta_1} &= \left\|\left(1-\frac{u_t^2}{v_t}\right)^2 \frac{1}{v_t^2}\left(\lambda_1 \frac{\partial v_t}{\partial \alpha} + \lambda_2 \frac{\partial v_t}{\partial \beta}\right)^2\right\|_{1+\eta_1} + o(1) \\
E\left|\nabla_t\right| &= E\left[\left|1-\frac{u_t^2}{v_t}\right| \frac{1}{v_t}\left(|\lambda_1| \frac{\partial v_t}{\partial \alpha} + |\lambda_2| \frac{\partial v_t}{\partial \beta}\right)\right] + o(1), \text{ as } t \to +\infty.
\end{aligned}$$

This yields

$$\begin{aligned}
\sum_{t=1}^n E\left[\eta_{nt}^2 \mathbb{1}_{|\eta_{nt}|>\varepsilon}\right] \leq\ & K\frac{1}{n\sqrt{n}} \sum_{t=1}^n \left\{\left(\left\|\left(1-\frac{u_t^2}{v_t}\right)^2 \frac{1}{v_t^2}\left(\lambda_1 \frac{\partial v_t}{\partial \alpha} + \lambda_2 \frac{\partial v_t}{\partial \beta}\right)^2\right\|_{1+\eta_1} + \right.\right. \\
& \left. E\left[\left(1-\frac{u_t^2}{v_t}\right) \frac{1}{v_t}\left(\lambda_1 \frac{\partial v_t}{\partial \alpha} + \lambda_2 \frac{\partial v_t}{\partial \beta}\right)\right]\right) \\
& \left. E\left[\left|1-\frac{u_t^2}{v_t}\right| \frac{1}{v_t}\left(|\lambda_1| \frac{\partial v_t}{\partial \alpha} + |\lambda_2| \frac{\partial v_t}{\partial \beta}\right)\right] + K_t^{(1)}\right\},
\end{aligned}$$

where $K_t^{(1)} = o(1)$ as $t \to +\infty$. Using the Cesàro lemma, Assumption **B6** and the stationarity of (v_t), $(\frac{\partial v_t}{\partial \alpha})$ and $(\frac{\partial v_t}{\partial \beta})$, we conclude and obtain the Lindeberg condition. Consequently, we have

$$\sum_{t=1}^n \eta_{nt} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, E\left[\left(1-\frac{u_1^2}{v_1}\right)^2 \frac{1}{v_1^2}\left(\lambda_1 \frac{\partial v_1}{\partial \alpha} + \lambda_2 \frac{\partial v_1}{\partial \beta}\right)^2\right]\right).$$

Then, using Lemma 4.6.14 and the Cramér-Wold device, we obtain (4.6.52).

We now prove that the matrix A_{α_0,β_0} is definite positive. If this matrix is not definite positive, then there exists $\lambda \in \mathbb{R}^2$ such that $\lambda' E\left[\left(1-\frac{u_1^2}{v_1}\right)^2 \frac{1}{v_1^2} \frac{\partial v_1}{\partial(\alpha,\beta)'} \frac{\partial v_1}{\partial(\alpha,\beta)}\right]$ and we have necessarily

$$\left(1-\frac{u_1^2}{v_1}\right) \frac{1}{v_1}\left(\lambda_1 \frac{\partial v_1}{\partial \alpha} + \lambda_2 \frac{\partial v_1}{\partial \beta}\right) = 0, \text{ a.s.}$$

As in the proof of Lemma 4.6.6, we found that this implies $\lambda_1 \frac{\partial v_t}{\partial \alpha} + \lambda_2 \frac{\partial v_t}{\partial \beta} = 0$ almost surely. We have

$$\begin{aligned}
\frac{\partial v_{t+1}}{\partial \alpha} &= \frac{u_t^2}{a_t} + \frac{\beta}{a_t}\frac{\partial v_t}{\partial \alpha} \\
\frac{\partial v_{t+1}}{\partial \beta} &= \frac{\alpha}{a_t}\frac{\partial v_t}{\partial \alpha} + \frac{\beta}{a_t}\frac{\partial v_t}{\partial \beta}.
\end{aligned}$$

Using the stationarity of $(\frac{\partial v_t}{\partial \alpha})$ and $(\frac{\partial v_t}{\partial \beta})$, we deduce

$$\lambda_1 u_t^2 = -\lambda_2 \alpha \frac{\partial v_t}{\partial \alpha}.$$

The left term of the last equation depends only on u_t, the right term depends on $\{u_i,\ i < t\}$, therefore we can conclude and we obtain $\lambda_1 = \lambda_2 = 0$ and consequently the matrix A_{α_0,β_0} is definite positive. $\square$

Lemma 4.6.16. *Let ϖ be an arbitrary compact subset of $(0, +\infty)$. Under the assumptions of Theorem 4.4.3, we have for $(i,j) \in \{2,3\}^2$*

$$\sup_{\omega \in \varpi} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\omega, \alpha_0, \beta_0) - C_{\alpha_0,\beta_0}(i-1, j-1) \right| \to 0, \ \textit{a.s. when } n \to +\infty. \qquad (4.6.53)$$

Proof. We have, for any $\theta \in \Theta$

$$\begin{aligned}
\left| \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} - D_t^{\theta_i,\theta_j} \right| \leq\ & \frac{h_t}{\sigma_t^2}\left(1 - u_t^2 \frac{h_t}{\sigma_t^2}\right)\left(\frac{1}{h_t}\frac{\partial^2 \sigma_t^2}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 v_t}{\partial \theta_i \partial \theta_j}\right) \\
& + \frac{\partial^2 v_t}{\partial \theta_i \partial \theta_j}\left\{\frac{h_t}{\sigma_t^2}\left(1 - u_t^2 \frac{h_t}{\sigma_t^2}\right) - \frac{1}{v_t}\left(1 - \frac{u_t^2}{v_t}\right)\right\} \\
& + \frac{h_t^2}{\sigma_t^4}\left(2u_t^2 \frac{h_t}{\sigma_t^2} - 1\right)\left\{\left(\frac{1}{h_t}\frac{\partial \sigma_t^2}{\partial \theta_i} - \frac{\partial v_t}{\partial \theta_i}\right)\left(\frac{1}{h_t}\frac{\partial \sigma_t^2}{\partial \theta_j} - \frac{\partial v_t}{\partial \theta_j}\right)\right. \\
& \left. + \left(\frac{1}{h_t}\frac{\partial \sigma_t^2}{\partial \theta_i} - \frac{\partial v_t}{\partial \theta_i}\right)\frac{\partial v_t}{\partial \theta_j}\frac{\partial v_t}{\partial \theta_i}\left(\frac{1}{h_t}\frac{\partial \sigma_t^2}{\partial \theta_j} - \frac{\partial v_t}{\partial \theta_j}\right)\right\} \\
& + \frac{\partial v_t}{\partial \theta_i}\frac{\partial v_t}{\partial \theta_j}\left\{\frac{h_t^2}{\sigma_t^4} 2u_t^2 \left(\frac{h_t}{\sigma_t^2} - \frac{1}{v_t}\right) + \left(2\frac{u_t^2}{v_t} - 1\right)\left(\frac{h_t^2}{\sigma_t^4} - \frac{1}{v_t^2}\right)\right\}.
\end{aligned}$$

With the arguments used to establish (4.6.28) and (4.6.45), we can also obtain

$$\sup_{\omega\in\varpi}\left|v_t(\alpha_0,\beta_0)-\frac{\sigma_t^2(\omega,\alpha_0,\beta_0)}{h_t}\right|\to 0,\quad \sup_{\omega\in\varpi}\left|\frac{\partial v_t(\alpha_0,\beta_0)}{\partial\theta_i}-\frac{1}{h_t}\frac{\partial\sigma_t^2(\omega,\alpha_0,\beta_0)}{\partial\theta_i}\right|\to 0$$
$$\sup_{\omega\in\varpi}\left|\frac{\partial^2 v_t(\alpha_0,\beta_0)}{\partial\theta_i\partial\theta_j}-\frac{1}{h_t}\frac{\partial^2\sigma_t^2(\omega,\alpha_0,\beta_0)}{\partial\theta_i\partial\theta_j}\right|\to 0,\text{ a.s. when } t\to+\infty.$$

Using the ergodicity of the process (v_t) and its derivatives, this yields (4.6.53).

□

Lemma 4.6.17. *Under the assumptions of Theorem 4.4.3, we have, for $(i,j,k)\in\{1,2,3\}^3$*

$$\sum_{t=1}^{+\infty}\sup_{\theta\in\Theta_\gamma}\left|\frac{\partial l_t(\theta)}{\partial\omega}\right|<+\infty,\ a.s. \tag{4.6.54}$$
$$\sum_{t=1}^{+\infty}\sup_{\theta\in\Theta_\gamma}\left\|\frac{\partial^2 l_t(\theta)}{\partial\omega\partial\theta}\right\|<+\infty,\ a.s. \tag{4.6.55}$$
$$\limsup_n\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}l_t(\theta)\right|<+\infty,\ a.s. \tag{4.6.56}$$

Proof. For any $\theta\in\Theta_\gamma$, we have

$$\frac{\partial l_t}{\partial\omega}=\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\omega}\left(1-u_t^2\frac{h_t}{\sigma_t^2}\right)=\left(1-u_t^2\frac{h_t}{\sigma_t^2}\right)\frac{1}{\sigma_t^2}\sum_{j=1}^{t}\beta^{j-1}.$$

Using Lemma 4.6.9 and the fact that Θ is a compact, we obtain

$$\sup_{\theta\in\Theta}\left|\frac{\partial l_t}{\partial\omega}\right|\le\left|1+u_t^2V_t\right|K\rho^t.$$

Since $\sum_{t=1}^{+\infty}\left|1+u_t^2V_t\right|K\rho^t$ has finite expectation, it is almost surely finite and we have (4.6.54). (4.6.55) can be obtained with the same arguments.

It remains to prove (4.6.56). For any $\theta\in\Theta_\gamma$, for $(i,j,k)\in\{2,3\}^3$, $\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}$ can be written as in the stationary case by (4.6.22). Using already used arguments, we obtain

$$\begin{aligned}\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} &= \left(2\frac{u_t^2}{v_t}-1\right)\frac{1}{v_t^2}\left(\frac{\partial v_t}{\partial\theta_i}\frac{\partial^2 v_t}{\partial\theta_j\partial\theta_k}+\frac{\partial v_t}{\partial\theta_j}\frac{\partial^2 v_t}{\partial\theta_i\partial\theta_k}+\frac{\partial v_t}{\partial\theta_k}\frac{\partial^2 v_t}{\partial\theta_i\partial\theta_j}\right)\\ &\quad+\left(1-\frac{u_t^2}{v_t}\right)\frac{1}{v_t}\frac{\partial^3 v_t}{\partial\theta_i\partial\theta_j\partial\theta_k}+\left(2-6\frac{u_t^2}{v_t}\right)\frac{1}{v_t^3}\frac{\partial v_t}{\partial\theta_i}\frac{\partial v_t}{\partial\theta_j}\frac{\partial v_t}{\partial\theta_k}+o(1),\text{ a.s.}\end{aligned}$$

We use the same arguments as those used to obtain (4.6.46) to prove that for any $y > 0$, for any $\theta \in \Theta^{(y)}$ and for any $(i, j, k) \in \{2, 3\}^3$ we have $\left\| \frac{\partial^2 v_t}{\partial \theta_i \partial \theta_j} \right\|_y < +\infty$ and $\left\| \frac{\partial^3 v_t}{\partial \theta_i \partial \theta_j \partial \theta_k} \right\|_y < +\infty$. Using the Hölder inequality, we obtain (4.6.56). □

We are now able to prove Theorem 4.4.3. As in the stationary case, we can write a Taylor expansion of the criterion $I_n(\theta)$ at θ_n, except that in this case, we do not have a consistency result for ω_n and the derivative of the criterion is not necessarily equal to zero at θ_n. For any $\omega_0 > 0$, we define $\theta_0 = (\omega, \alpha_0, \beta_0)$ and we write the Taylor expansion at $\theta = \theta_0$ and we keep only the two last lines of this equality, we have

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial}{\partial \omega} l_t(\theta_n) \\ 0 \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial}{\partial \theta} l_t(\theta_0) + \mathcal{J}_n \sqrt{n} \left(\theta_n - \theta_0\right),$$

where $\mathcal{J}_n$ is a 3×3 matrix whose element have the form

$$\mathcal{J}_n = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} l_t(\theta_i^*),$$

where θ_i^* is between θ_n and θ_0. Since $\theta_0 \in \Theta_\gamma$, for n large enough, we have $\theta_i^* \in \Theta_\gamma$. Using Assumption **B2** and Lemma 4.6.17, we obtain, for $i \in \{2, 3\}$

$$\mathcal{J}_n(i, 1) \sqrt{n} \left(\omega_n - \omega_0\right) \to 0, \text{ a.s.}$$

For $(i, j) \in \{1, 2\}^2$, using Lemma 4.6.16, Lemma 4.6.17 and another Taylor expansion of the function

$$(\alpha, \beta) \mapsto \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\omega_i^*, \alpha, \beta),$$

we obtain

$$\mathcal{J}_n(1 + i, 1 + j) \to C_{\alpha_0, \beta_0}(i, j), \text{ a.s. when } n \to +\infty. \tag{4.6.57}$$

Bibliography

E. Andreou and E. Ghysels. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics*, 17(5):579–600, 2002.

L. Bauwens, A. Preminger, and J.V.K. Rombouts. Theory and inference for a Markov switching GARCH model. *The Econometrics Journal*, 13(2):218–244, 2010.

P. Billingsley. *Probability and Measure.* John Wiley & Sons, 1995. Third Edition.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

J. Cai. A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, pages 309–316, 1994.

I. Domowitz and H. White. Misspecified models with dependent observations. *Journal of Econometrics*, 20(1):35–58, 1982.

M.J. Dueker. Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business & Economic Statistics*, pages 26–34, 1997.

R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, pages 987–1007, 1982.

C. Francq and J.M. Zakoïan. The L2-structures of standard and switching-regime GARCH models. *Stochastic processes and their applications*, 115(9):1557–1582, 2005.

C. Francq and J.M. Zakoïan. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–638, 2004.

C. Francq and J.M. Zakoïan. Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica*, 80(2):821–861, 2012.

C. Francq, M. Roussignol, and J.M. Zakoian. Conditional heteroskedasticity driven by hidden Markov chains. *Journal of Time Series Analysis*, 22(2):197–220, 2001.

Sylvia Fruhwirth-Schnatter and Sylvia Kaufmann. Bayesian Analysis of Switching ARCH Models. *Journal of Time Series Analysis*, 37(4), 2002.

C. Gourieroux, A. Monfort, and A. Trognon. Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700, 1984.

S.F. Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62, 1996.

J.D. Hamilton and R. Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1-2):307–333, 1994.

J.S. Henneke, S.T. Rachev, F.J. Fabozzi, and M. Nikolov. MCMC-based estimation of Markov Switching ARMA–GARCH models. *Applied Economics*, 43(3):259–271, 2011.

E. Hillebrand. Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129 (1):121–138, 2005.

A.T. Jensen and T. Lange. On convergence of the QMLE for misspecified GARCH models. *Journal of Time Series Econometrics*, 2(1):1–29, 2010.

C. Klüppelberg, A. Lindner, and R. Maller. A continuous-time GARCH process driven by a Lévy process: Stationarity and second-order behaviour. *Journal of Applied Probability*, 41(3): 601–622, 2004.

J.C. Liu. Stationarity of a Markov-switching GARCH model. *Journal of Financial Econometrics*, 4(4):573–593, 2006.

L. Mevel and L. Finesso. Asymptotical statistics of misspecified hidden Markov models. *Automatic Control, IEEE Transactions on*, 49(7):1123–1132, 2004.

T. Mikosch and C. Stărică. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, pages 378–390, 2004.

Y. Ogata. Maximum likelihood estimates of incorrect Markov models for time series and the derivation of AIC. *Journal of Applied Probability*, pages 59–72, 1980.

K. Tanaka and K. Maekawa. The sampling distributions of the predictor for an autoregressive model under misspecifications. *Journal of Econometrics*, 25(3):327–351, 1984.

H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, pages 1–25, 1982.